Technische Universität Dortmund

Department of Statistics

OPTIMAL ESTIMATION OF NONLINEAR FUNCTIONS OF DISTRIBUTION PARAMETERS

Thesis submitted in fulfillment of the requirements for the degree of Doktor der Naturwissenschaften at the Technische Universität Dortmund

By

Ana Moya

Supervisor: Prof. Dr. Trenkler

Dortmund, August 2009

Acknowledgements

I would like to take this opportunity to express my thanks to my thesis advisor Prof. Dr. Götz Trenkler, who guided me meticulously through the research work. Not only he guided me on scientific issues, but he has been also very understanding and cooperative.

I also express my deepest gratitude to the “Deutsche Forschungsgemeinschaft” for financial support and for providing me with an excellent and friendly working environment in the Graduiertenkolleg ”Statistische Modellbildung”.

Special mention is also due to Eva Brune, Lars Willmes, Dr. Matthias Arnold and Antonio Sucre who supported me through many important aspects of this work.

I would also like to thank to my family members in Venezuela, my friends in Dortmund and colleagues at NuTech Solutions for their great inspiration, encouragement and support, all of which I will cherish throughout my life.

Finally, I thank Almighty God, for all the grace and blessing that he has showered on me to reach this stage of my academic and personal life.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 General Motivation

In statistics, inference on parameters of a given probability distribution is usually made by means of point estimation, based on observed data from the distribution under study.

In several fields of statistical practice though, it is of interest to estimate functions of parameters instead of the parameters themselves. Nonlinear functions, see Definition 16, are useful for describing the properties of a certain population or even for comparing the properties of two different populations.

The main motivation for this type of approach comes from Casella and Berger (1990, p. 330) who investigated the estimation of two specific nonlinear functions; the inverse of the mean, see Definition 11, and the ratio of two means, see Definition 9. For the estimation of such functions in the aforementioned work a first-order approximation of the arithmetical mean, see Definition 7, was used.

This work will focus on sufficiently smooth functions which are functions that have continuous derivatives up to some desired order, see Definition 1.

Neudecker and Trenkler (2002) show that in good approximation such a function can be written as a sum of a linear and a quadratic form, see Definition 5. Similarly, Taylor Series provide a means of approximating a function through polynomials, see Definition 3.

The fundamental aim of this work is to introduce a new method for handling the statistical inference of functions of distribution parameters $\boldsymbol{\theta}$ or of random variables. Due to its complex nature, this work will focus specifically on nonlinear, sufficiently smooth (NLSS) functions, see Definition 2. For a review of the underlying theoretical principles, ideas and methods of point estimation, reference is made to standard books by Lehmann (1983) and Zacks (1971), the former containing a more intuitive and applied approach towards this topic.

For the estimation of transformations or NLSS functions $f(\boldsymbol{\theta})$ of distribution parameters $\boldsymbol{\theta}$ under simple random schemes, the usual choices are: either to calculate the transformation of the estimator of $\boldsymbol{\theta}$, or to calculate the estimator of the transformed data. In the first approach the estimation is made before the transformation and in the second the estimation is made after the transformation, i.e. the function is applied on the estimator in the first case and on the data in the second. This has influence on the bias and the variance of the respective estimators, which will be shown in Section 2.2.

In comparative analysis, effects are commonly expressed as ratios. In particular, in bio-assays the calculation of the relative potency requires the estimation of the ratio of two normal means, see Definition 9. The first work in this context, and the extension to multiple regression, was presented by Finney (1971). In Section 2.1 former research on this topic is presented.

In statistical practice, sometimes the data need to be transformed; such transformations are frequently used to stabilise the variance and/or to produce linearity or additivity. Reduction in non-normality might also require such transformations, especially a logarithmical transformation, see Section 6.2.2. The ratio of two lognormal means has been widely discussed in the bio-statistic literature, where different estimators have been deduced. Estimations of this kind are commonly used in studies of equivalence of treatments, e.g. in the bio-equivalence analysis.

In general, statistical inference is more intricate for a ratio of parameters than for linear combinations of them. The task of making inferences about the ratio of two normal means can be addressed within at least two different scenarios. First, if the corresponding variances are equal, a simple solution can be obtained using the Bayesian approach, see Bernardo (1977) and Bernardo and Ramón (1998). Second, if the assumption of equal variances cannot be sustained, the frequentist approach can offer approximate (asymptotic) answers.

Estimation of the inverse of the population mean, see Definition 11, is used in many situations, for instance in Econometrics and Biological Sciences, see Zellner (1978), who presented a Minimum Expected Loss (MELO) estimator. The MELO and Maximum Likelihood (ML) estimators, see Definition 24, have very different finite sample properties; but as the sample size becomes larger, their large sample distributions become identical.

1.2 Problem Statement

Most published methods concerning the estimation of nonlinear functions, though, are asymptotic in nature as well as based on the assumption of normal distribution and that the random variables involved are not correlated. Therefore, the general application of those methods of inference in small-sample data analysis, particularly for data from biological and medical experiments, would be nearly impossible. For instance, existing methods for comparing the means of two independent skewed lognormal distributions by means of ratios have been shown not to perform well in a range of small-sample settings such as a bio-availability study, see Crow (1977) and Shaban (1981).

The new inference method, to be presented in this work, is not based on any assumption of any type of distribution or of any data structure. In this work different simulations will be carried out in order to observe how the new method works under different distribution assumptions and sample sizes. This method can even be used for comparing the parameters of two correlated distributions. To enable this comparison a new approach for generating correlated random variables will also be developed.

The inference methods presented in literature have several good properties, which will be discussed in Section 2. One inherent disadvantage of one of these methods is the non-existence of higher order moments. This has significant implications in different data analysis situations. In particular, this problem appears when the inverse of the mean, see Definition 11, is the function of interest. Due to the invariance property, Zacks (1971), p. 223, the inverse of the sample mean is the maximum likelihood estimator of the inverse of the population mean under normality. Such an estimator is biased and has no second or higher order moments, which is also the case with quite a few other distributions. The problem of non-existence of moments is explained in Appendix A.6.

The first and second moments of the new inference method will be approximated, on the basis of a linear plus quadratic function, as presented in Neudecker and Trenkler (2005a). Thus, the problem of non-existence of higher order moments does not appear in this new approach. For the calculation of these moments for correlated random variables, a new approach based on Kleffe and Rao (1988) will also be developed.

1.3 Structure

This work is organised as follows:

In Section 2, NLSS functions are defined, see Definition 2. The general procedure used for the estimation of such a function is also presented as well as previous investigations made into this subject.

As mentioned in Section 1.1, the main motivation for the estimation of NLSS functions comes from Casella and Berger (1990). Their general procedure will also be presented in Section 2, as well as the mathematical foundations of Taylor Series, necessary for the approximation of these functions.

The estimation of NLSS functions, such as $f(\boldsymbol{\theta})$, with $\boldsymbol{\theta}$ a vector of parameters of a given distribution, is presented in Section 3. Two inference methods, the transformation of the estimator of the parameter involved, and the estimator of the transformed data, as well as their basic properties are also illustrated in this section.

Approximated means and variances of the aforementioned estimators, as well as an approximation of the covariance between them, are presented in this section.

In Section 4 an unbiased estimator for approximated NLSS functions $f(\boldsymbol{\theta})$, derived by using the generalised Jackknife approach, will be presented.

In Section 5, it will be shown that the variance of the generalised Jackknife estimator can be minimised with small bias. In the same section it will be explained how this minimisation can be made and at the same time the most important properties of the resulting estimator will be presented. The resulting estimator will have minimal Mean Squared Error (MSE).

In Section 6 estimation of some specific NLSS functions of parameters of distributions will be investigated. As first NLSS function the estimation of the ratio of means is considered, for this function different estimators from literature will be compared, by means of their approximated MSEs and of simulations, with the generalised Jackknife estimator and with the minimal MSE deduced in this work. Other functions to be investigated are the inverse of the mean and the odds used for the calculation of the odds ratio.

Due to the nature of the new approach presented, several unsolved problems arise. In Section 7 such problems and challenges are presented.

Chapter 2

Estimation of NLSS Functions

Contents

In this section, the most significant results concerning the estimation of NLSS functions presented in literature will be described.

The relevant definitions and general procedure used to develop a new inference method of NLSS functions are also presented in this section.

2.1 Background and Literature Review

One of the first works on inference for ratios of normal means with application to bioassays was published by Fieller (1944). Based on the fiducial argument Fieller and Creasy (1954) proposed different solutions to the problem of building a confidence interval for the ratio of means. This solution was controversial and prompted several discussions because conventional methods had no principles for dealing with it. It is a prime example of an estimation problem for which only Bayesian methods provide the technical requirements to solve it. On the other hand, the main difficulty from the frequentist point of view is that, with probability one, the expected longitude of any confidence interval for the ratio of two normal means is infinite (Gleser and Hwang (1987)). In literature, this problem has been approached from a Bayesian point of view. One of the first works in which this can be seen is that presented by Bernardo (1977), who obtained the final distribution for the case where the sample sizes and the variances are equal.

Crow (1977) derived a Minimum Variance Unbiased Estimator (MVUE) for the ratio of means of two independent gamma distributed variables, with known shape as well as for two lognormal distributed variables, with equal and unequal shape parameters.

Zellner (1978) recommended analysing the problem of estimating the inverse of the mean and the ratio of means from a Bayesian point of view. For this approach it is assumed that the observations are normally and independently distributed, each with mean μ and variance σ^2, and both are unknown. He developed the Minimum Expected Loss (MELO) estimator of the inverse of the mean.

Srivastava and Bhatnagar (1981) have erroneously stated that the Maximum Likelihood estimator of the inverse of the mean has no finite moments. This statement holds only for the second and higher order moments, as in Voinov (1985) the first moment of the Maximum Likelihood estimator of the inverse of the mean is presented, see Appendix B.2.3. Based on the approach presented by Zellner (1978), Srivastava and Bhatnagar (1981) presented a class of estimators which are free from the limitation of non-existence of moments. They derived exact expressions for the first two moments in the case of normal populations and they also proposed large sample approximations for non-normally distributed populations.

Shaban (1981) obtained estimators for the ratio of means of two independent lognormal distributed variables which are generally of smaller mean squared error than both the Maximum Likelihood (ML) and the MVUE as given in Crow (1977).

Barndorff-Nielsen (1983) presented one of the first works where the problem in question is analysed using the distribution of the Maximum Likelihood estimator.

Buonaccorsi (1985) considered the problem of estimating the ratio of two linear combinations of the vector of parameters in the general linear model. He also discussed the non-existence of an unbiased estimator under normal errors. In his work the author, as well as Fieller (1932) and Srivastava and Bhatnagar (1981), declared that the first and higher order moment of the inverse of the arithmetical mean does not exist.

Voinov (1985) has derived unbiased estimators of powers of the inverse of the normal population mean, μ. In the same work, it is shown that such an estimator is useful in the solution of several problems in experimental nuclear physics.

Casella and Berger (1990) advocated using a first-order approximation for the estimation the inverse of the mean and the ratio of means.

Tiwari and Elston (1999) approximated the variance of a function of random variables by using a second degree Taylor series expansion, and demonstrated the increased accuracy this second degree approximation gives over the usual Delta method by using some examples from genetics.

Rao (2002) highlighted two estimators for the ratio of means. He pointed out that under simple random sampling scheme, the usual choices for the estimation of the ratio of means are (i) a (single) ratio of sample means or (ii) the mean of (n) ratios. He also reported that both estimators are biased for the ratio of means. For the estimation of this function he considered a class of Symmetrized Des Raj (SDR) strategies and looked for a choice of a model-optimum estimator when design-unbiasedness is not demanded, among those utilising "mean of ratios" and "ratio of means", as he denominated these estimators.

Troschke (2002) analysed forecasts for the future development of key variables in the field under consideration. Specifically, he studied the case where the decision-maker has the problem of having more than one forecast for the variables of interest, then instead of selecting one of the forecasts it is a successful strategy and common practice to combine the individual forecasts. His predominant approach was to concentrate on one target variable at a time and to perform a linear combination of the forecasts for that variable. In his work, this standard univariate linear combination approach was improved in two respects: Firstly, a linear plus quadratic set-up for the combination of univariate forecasts was introduced as a non-linear combination alternative (univariate linear plus quadratic combination). Similar to a higher order Taylor approximation it may result in more accurate combined prediction. Secondly, several target variables are considered at the same time, and forecasts for such vector valued variables were combined linearly (multivariate linear combination). Thus, additional information was exploited by taking the interactions between the components of target vector and forecasts into account. For each approach the mean square prediction error optimal combination parameters were derived. Finally, the new approaches were investigated numerically in terms of their potential, their empirical performance and their results in a simulation study.

Neudecker and Trenkler (2002) considered the problem of estimating a function of the multivariate mean. They assumed that in good approximation this function can be written as a sum of a linear and a quadratic form. For the estimators presented in their work they obtain mean and variance, when the populations are independent. The same problem was also considered by Frauendorf, Neudecker and Trenkler (2005). In their work, it is shown that two naive estimators turn out to be biased. Using a generalized jackknife procedure they constructed an unbiased estimator of this function as a reasonable alternative. Variances of the three estimators were calculated for the general and the normal case.

Neudecker and Trenkler (2005a) considered the problem of approximating the variance of nonlinear functions of random variables using a second degree Taylor series expansion. In contrast to Tiwari and Elston (1999), their approach also uses the covariances between the random variables to obtain a better approximation.

Qiao et al. (2006) addressed the problem of estimating the ratio of the means of independent normal variables in agricultural research. In the first part of their research, the authors examined the distributional properties of the ratio of independent normal variables, both theoretically and using simulation. In the second part of their research, they evaluated the relative merits of two common estimators of the ratio of the means of independent normal variables in agricultural research, an arithmetic average or ratio of means and a weighted average or mean of ratios, via simulation experiments using normal distributions. They gave a condition under which the mean of the ratio of two independent normal variables appears to exist and that can be used to evaluate the suitability of both estimators. In this work it is mentioned that the development of a satisfactory estimator of the ratio when the involved random variables are dependent remains an area for future research.

Friedrich et al. (2008) investigated the ratio of means method as an alternative to mean differences for analyzing continuous outcome variables in meta-analysis. Meta-analysis of continuous outcomes traditionally uses mean difference or standardized mean difference, i.e. mean difference in pooled standard deviation units. They pointed out that both the standardized mean difference and the ratio of means allow pooling of outcomes expressed in different units and comparisons of effect sizes across interventions, but the ratio of means interpretation does not require knowledge of the pooled standard deviation, a quantity generally unknown to clinicians.

2.2 Definitions and Methods

The fundamental aim of this work is to introduce a new approach for handling the statistical inference of NLSS functions by means of approximations. Before the general procedure of this approach is explained it is essential; a) to introduce necessary definitions as well as clarifying remarks, b) to clarify how NLSS functions are defined and c) to illustrate the approach to be used for the approximation of those functions. For concepts related to calculus and probability theory refer to Gradshteyn and Ryzhik (2000), Shorack (2000) as well as Mood, Graybill and Boes (1974).

The purpose of this section is to introduce the general definitions necessary for the development of a new inference method of NLSS functions. It is primarily a "definitions-and-their-understanding" section.

2.2.1 Types of Functions

In this section as well as in Appendix A.1.1 the most important concepts necessary to clarify how NLSS functions are defined, are presented.

For the definition of NLSS functions the concept of *smooth functions* has to be introduced. This kind of functions is presented in the following definition.

Definition 1 (Smooth Functions)

Let $\mathfrak{B}$, $\mathfrak{C} \subseteq \mathbb{R}$, where $\mathfrak{B}$ and $\mathfrak{C}$ are unions of open intervals. Furthermore, let K be a non-negative integer and $f^{(i)}$ the i-th derivative *of the function f. A function $f : \mathfrak{B} \to \mathfrak{C}$ is said to be of class C^K if the derivatives $f^{(1)}, f^{(2)}, ..., f^{(K)}$ exist and are continuous. In the same way, the function f is said to be of class C^∞, or* smooth, *if it has derivatives of all orders.*

After smooth functions have been defined, the kind of function to be analysed in this work is presented in the following definition.

Definition 2 (Nonlinear, Sufficiently Smooth (NLSS) Functions)

For a given order K, a function f, say $f : \mathfrak{B} \to \mathfrak{C}$, is said to be a NLSS function of order K when it is nonlinear and of at least class C^K.

For additional definitions refer to Appendix A.1.1.

2.2.2 Types of Approximations

As mentioned in Section 1.1 the main motivation for estimating functions by approximations comes from Casella and Berger (1990, p. 330). They investigated the estimation of the ratio of two means, see Definition 9, and the inverse of the mean, see Definition 11. In their approach a Taylor Series expansion of order one was used for the estimation of the aforementioned functions.

The mathematical foundations, necessary for the approximation of the functions under study, are presented in this section.

Definition 3 (Taylor Series Expansion)

The Taylor Series expansion of a real or complex function $f(x)$ that is infinitely often differentiable in a neighborhood of a real or complex number x_0 is defined as:

$$f(x) = \sum_{d=0}^{\infty} \frac{f^{(d)}(x_0)}{d!}(x - x_0)^d,$$

where (d) represents the order of the derivatives of f.

Suppose a function $f(x)$ of at least class C^K is given, the Taylor Series expansion of order K of $f(x)$ in a neighborhood of x_0 is defined as:

$$f(x) = \sum_{d=0}^{K} \frac{f^{(d)}(x_0)}{d!}(x - x_0)^d + Error,$$

where each term higher than K is contained in the error term, denoted by Error, and it is assumed to be neglectable.

For further details see Mera (1992).

Remark 2.1 (Approach according to Casella and Berger (1990))

Suppose $\bar{x}$ is the arithmetical mean of a random sample, see Definition 7. If a function $f(\mu)$ has to be estimated, using a Taylor Series expansion of order $K = 1$ Casella and Berger (1990) presented the following approximated estimator for $f(\mu)$:

$$f(\bar{x}) = f(\mu) + f^{(1)}(\mu)(\bar{x} - \mu).$$

They also stated that the estimator $f(\bar{x})$ has the following approximated mean and variance:

$$\mathrm{E}[f(\bar{x})] \approx f(\mu), \textit{ and } \mathrm{var}(f(\bar{x})) \approx [f^{(1)}(\mu)]^2 \mathrm{var}(\bar{x}),$$

where $f^{(1)}(\mu)$ represents the first order derivative of $f(\mu)$.

Stimulated by the approach presented in Casella and Berger (1990) and Taylor Series expansion, in this work, it is assumed that a function $f(x)$ can be sufficiently good approximated by a Taylor Series expansion of order $K = 2$.

This approximation is given as follows:

$$f(x) \approx \underbrace{f(x_0) + f^{(1)}(x_0)(x - x_0) + \frac{1}{2!} f^{(2)}(x_0)(x - x_0)^2}_{f^{Taylor}(x)}, \qquad (2.2.1)$$

where $f^{(1)}(x_0)$ and $f^{(2)}(x_0)$ represent the first and second order derivatives of $f(x_0)$, respectively. For this approach, the function $f(x_0)$ is necessarily of at least class C^2, see Definition 1. The index *Taylor* stands for representing the Taylor approximation.

Expression 2.2.1 can be expanded to the p-dimensional case, i.e. the function to be approximated is not applied on a real value x, but is applied on a vector $\boldsymbol{x}$. In such a case it is necessary to use multidimensional versions of the first and second order derivatives of $f(\boldsymbol{x}_0)$.

In this work it is assumed that the gradient vector and the Hessian matrix, containing the first and second order partial derivatives with respect to the vector $\boldsymbol{x} = (x_1, \ldots, x_p)'$, can be used as multidimensional versions of the first and second order derivatives, $f^{(1)}$ and $f^{(2)}$. These concepts are presented in the following definition.

Definition 4 (Gradient Vector and Hessian Matrix)

Let $f : \mathfrak{B}^p \rightarrow \mathfrak{C}$ be a function defined on the p-dimensional Euclidean space, where $\mathfrak{B}$ and $\mathfrak{C}$ are cartesian products of unions of open intervals. The gradient of a function $f(\boldsymbol{x})$ with respect to the vector $\boldsymbol{x} = (x_1, \ldots, x_p)'$ is defined as a vector whose components are the partial derivatives of the function $f(\boldsymbol{x})$. It is defined as follows:

$$f^{(1)}(\boldsymbol{x}) = \left(\frac{\partial f(\boldsymbol{x})}{\partial x_1}, \ldots, \frac{\partial f(\boldsymbol{x})}{\partial x_p} \right)'.$$

If all second partial derivatives of $f(\boldsymbol{x})$ exist, then the square matrix of second-order partial derivatives of this function, the Hessian matrix, *is defined as:*

$$f^{(2)}(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1^2} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1\,\partial x_2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1\,\partial x_p} \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2\,\partial x_1} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2\,\partial x_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_p\,\partial x_1} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_p\,\partial x_2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_p^2} \end{pmatrix}.$$

The second partial derivatives of $f(\boldsymbol{x})$ with respect to two different variables are the off-diagonal entries in the Hessian. If they are all continuous, then the Hessian of $f(\boldsymbol{x})$ is symmetric.

For example, for a two-dimensional vector $\boldsymbol{x}$ symmetry is represented by:

$$\frac{\partial}{\partial x_1}\left(\frac{\partial f(\boldsymbol{x})}{\partial x_2}\right) = \frac{\partial}{\partial x_2}\left(\frac{\partial f(\boldsymbol{x})}{\partial x_1}\right) \Rightarrow \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1\,\partial x_2} = \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2\,\partial x_1}.$$

For further details see Gradshteyn and Ryzhik (2000).

Following expression 2.2.1 and Definition 4 in this work, it is assumed that a function $f(\boldsymbol{x})$ can be sufficiently good approximated in a neighborhood of the vector $\boldsymbol{x}_0$ by the following Taylor Series expansion of order $K = 2$:

$$f^{Taylor}(\boldsymbol{x}) = f(\boldsymbol{x}_0) + [f^{(1)}(\boldsymbol{x}_0)]'(\boldsymbol{x} - \boldsymbol{x}_0) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_0)' f^{(2)}(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0), \qquad (2.2.2)$$

where $[f^{(1)}(\boldsymbol{x}_0)]'$ represents the transpose of the gradient vector and $f^{(2)}(\boldsymbol{x}_0)$ the Hessian matrix of the function $f(\boldsymbol{x}_0)$.

Stimulated by Taylor Series expansions, Troschke (2002) investigated the *linear plus quadratic functions* approach aiming to find a combined forecast for a scalar random variable from several individual forecasts for that variable. This approach is presented in the following definition.

Definition 5 (Class of Linear plus Quadratic Functions)

Let a_0 be a constant, $\boldsymbol{a}$ a non-stochastic vector and $\boldsymbol{A}$ a non-stochastic symmetric matrix, then the class of linear plus quadratic functions is defined as:

$$\Theta = \left\{ a_0 + \boldsymbol{a}'\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} | a_0 \in \mathbb{R}, \boldsymbol{a} \in \mathbb{R}^p, \boldsymbol{A} \in \mathcal{M}_{p\times p} \right\},$$

where $\boldsymbol{a}'\boldsymbol{x}$ and $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ represent the linear and quadratic forms of this class, respectively. $\mathcal{M}_{p\times p}$ stands for the set of all symmetric $p \times p$ matrices.

Like Troschke (2002), in this work, also stimulated by Taylor Series expansions it is assumed that the function $f(\boldsymbol{x})$ may be approximated by the following linear plus quadratic function in $\boldsymbol{x}$:

$$f^{Poly}(\boldsymbol{x}) = a_0 + \boldsymbol{a}'\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}, \tag{2.2.3}$$

where a_0 is a constant, $\boldsymbol{a} = f^{(1)}(\boldsymbol{x})$ represents the gradient vector and $\boldsymbol{A} = \frac{1}{2}f^{(2)}(\boldsymbol{x})$ the Hessian matrix of the function $f(\boldsymbol{x})$ divided by two, see Definition 4. The index $Poly$ stands for representing the approximation of a function by a linear plus quadratic form, i.e. by a second order polynomial.

2.2.3 Estimation Approach

At this point the approximation of NLSS functions by means of Taylor Series expansions has been introduced. Now, the same approximation approach will be used for estimating NLSS functions of a given vector of interest based on *sample information.*

Before the general estimation approach is described, it is necessary to introduce the way how the sample information will be represented in this work. It will be represented in form of a matrix which will be called the *sample matrix* and is presented in the following definition.

Definition 6 (Sample Matrix)

Suppose p random variables $X_1, X_2, \dots, X_p$ with n observations taken on each of the variables are given. The corresponding $n \times p$ sample matrix is denoted by $\boldsymbol{X}$ and defined as:

$$\boldsymbol{X} = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{pmatrix},$$

For $i = 1, \cdots, n$ and for $j = 1, \cdots, p$, $\boldsymbol{X}$ can also be represented as:

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_n \end{pmatrix}, \text{ with } \boldsymbol{x}_i = \begin{pmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{ip} \end{pmatrix}', \text{ or as: } \boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_p \end{pmatrix}', \text{ with } \boldsymbol{x}_j = \begin{pmatrix} X_{1j} \\ X_{2j} \\ \vdots \\ X_{nj} \end{pmatrix},$$

for $i = 1, \cdots, n$ and for $j = 1, \cdots, p$.

Moreover, $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n$ represents a realisation of an independent, identically distributed (i.i.d.) random sample from a p–dimensional distribution, with mean $\boldsymbol{\mu} = E[\boldsymbol{x}_i]$ and covariance matrix $\boldsymbol{\Sigma} = E[(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})']$.

Definition 7 (Arithmetic Mean)

The arithmetic mean *of the j-th random variable in $\boldsymbol{X}$, denoted by $\bar{x}_j$, is defined as:*

$\bar{x}_j = \frac{1}{n} \sum_{i=1}^{n} X_{ij}$, $j = 1, \dots, p$.

The resulting p-vector of means is $\bar{\boldsymbol{x}} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p)'$, which alternatively can be written as $\bar{\boldsymbol{x}} = \frac{1}{n} \boldsymbol{X}' \mathbf{1}_n$, where $\mathbf{1}_n$ is the n-vector of ones.

Suppose a NLSS function $f(\boldsymbol{\theta})$ of a p-dimensional parameter vector $\boldsymbol{\theta}$ has to be estimated. Furthermore, suppose that an estimator of the vector $\boldsymbol{\theta}$ is given. For the estimation of NLSS function f of a given vector, say $\boldsymbol{\theta}$, under simple random schemes, the usual choices are either to apply the function on the estimator of $\boldsymbol{\theta}$ or to calculate the estimator of the transformed data. These estimation approaches are presented in the following definition.

Definition 8 (Estimation Approaches)

Let $\mathfrak{B}$, $\mathfrak{C} \subseteq \mathbb{R}$ and $\boldsymbol{X}$ be given as in Definition 6. Moreover, consider the following functions: $F : \mathfrak{B}^{n\times p} \to \mathfrak{C}^n$, $\hat{G} : \mathfrak{B}^{n\times p} \to \mathfrak{C}^p$, $f : \mathfrak{B}^p \to \mathfrak{C}$, $\hat{g} : \mathfrak{B}^n \to \mathfrak{C}$, with $\hat{G}$ and $\hat{g}$ representing estimators obtained from the sample information. Now, suppose a NLSS function f has to be estimated.

The first estimation approach is defined as follows:

$$T_1 = f(\hat{G}(\boldsymbol{X})) = f \begin{pmatrix} \hat{g}(\boldsymbol{x}_1) \\ \hat{g}(\boldsymbol{x}_2) \\ \vdots \\ \hat{g}(\boldsymbol{x}_p) \end{pmatrix}' = f\left(\hat{g}(\boldsymbol{x}_1), \hat{g}(\boldsymbol{x}_2), \cdots, \hat{g}(\boldsymbol{x}_p)\right), \text{ where } \boldsymbol{x}_1, \cdots, \boldsymbol{x}_p \in \mathfrak{B}^n.$$

The second estimation approach is defined as follows:

$$T_2 = \hat{g}(F(\boldsymbol{X})) = \hat{g} \begin{pmatrix} f(\boldsymbol{x}_1) \\ f(\boldsymbol{x}_2) \\ \vdots \\ f(\boldsymbol{x}_n) \end{pmatrix} = \hat{g}(f(\boldsymbol{x}_i), i = 1, \cdots, n), \text{ where } \boldsymbol{x}_1, \cdots, \boldsymbol{x}_n \in \mathfrak{B}^p.$$

In this approach $\mathfrak{B}$ and $\mathfrak{C}$ are defined as the *base set*.

In the first approach the estimation $\hat{G}$ is made before the function f is applied and in the second the estimation $\hat{g}$ is made after the application of the function f, i.e. the function f is applied on the estimator $\hat{G}(\boldsymbol{X})$ in the first case and on each row of the $n \times p$ sample matrix $\boldsymbol{X}$ in the second. This has influence on the bias and the variance of the respective estimators, which will be shown in Section 2.2.

Historically, $f(\hat{G}(\boldsymbol{X}))$ and $\hat{g}(F(\boldsymbol{X}))$ have been of interest even in the classical inference from infinite populations, see Rao (2002). Unfortunately, most of the authors have worked on the estimation of NLSS functions by considering $\boldsymbol{X}$ to be normally distributed. In this work both estimation approaches will be compared through a simulation study with different underlying distributions. These estimation approaches are the building blocks of the general estimation procedure to be developed in this work.

Tiwari and Elston (1999) approximated the variance of a function of random variables by using a second degree Taylor series expansion, and demonstrated the increased accuracy that this second degree approximation gives over the usual Delta method by using some examples from genetics. In the same respect, Neudecker and Trenkler (2005a) considered the problem of approximating the variance of a nonlinear function of random variables on the basis of a second degree Taylor series expansion. In contrast to the result achieved by Tiwari and Elston (1999), their approach in addition uses the covariances between the random variables to obtain a better approximation. This approach with its notation from Neudecker and Trenkler (2005a) is presented in the following remark.

Remark 2.2 (Approach by Neudecker and Trenkler (2005a))

As in Tiwari and Elston (1999), Neudecker and Trenkler (2005a) consider a scalar function f *of the random vector* $\mathbf{y} = (y_1, y_2, \cdots, y_m)'$ *with* $\mathrm{E}[\mathbf{y}] = \boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_m)'$ *and* $\boldsymbol{\Sigma} = \mathrm{D}(\mathbf{y}) = \mathrm{E}[(\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})']$ *denoting the expectation vector and the dispersion matrix, respectively, of* $\mathbf{y}$. *They assume that the first and second partial derivatives of* f *with respect to each* y_i $(i = 1, \cdots, m)$ *exist in an open neighborhood of* $\boldsymbol{\mu}$.

Let $\mathbf{y} = \boldsymbol{\mu} + \triangle\mathbf{y}$ *and* $f(\mathbf{y}) = f(\boldsymbol{\mu}) + \triangle f(\mathbf{y})$. *Using Taylor's formula they get the following approximation:*

$$\begin{aligned} f(\mathbf{y}) &\approx f(\boldsymbol{\mu}) + \frac{\partial f(\mathbf{y})}{\partial \mathbf{y}'} \,|_{\mathbf{y}=\boldsymbol{\mu}}\, d\mathbf{y} + \frac{1}{2}(d\mathbf{y})' \frac{\partial f^2(\mathbf{y})}{\partial \mathbf{y} \partial \mathbf{y}'} \,|_{\mathbf{y}=\boldsymbol{\mu}}\, d\mathbf{y} \\ &= f(\boldsymbol{\mu}) + \boldsymbol{a}' d\mathbf{y} + \frac{1}{2}(d\mathbf{y})' \boldsymbol{A} d\mathbf{y}. \end{aligned}$$

Since $\mathrm{E}[(d\mathbf{y})(d\mathbf{y})'] = \mathrm{D}(\mathbf{y}) = \boldsymbol{\Sigma}$ *and* $\mathrm{E}[(d\mathbf{y})] = 0$, *they get:*

$$\mathrm{E}[f(\mathbf{y})] \approx f(\boldsymbol{\mu}) + \frac{1}{2} tr(\boldsymbol{A}\boldsymbol{\Sigma})$$

$$\begin{aligned} \mathrm{D}(f(\mathbf{y})) &\approx \mathrm{D}\left(\boldsymbol{a}'(d\mathbf{y}) + \frac{1}{2}(d\mathbf{y})'\boldsymbol{A} d\mathbf{y}\right) \\ &= \mathrm{D}(\boldsymbol{a}'(d\mathbf{y})) + \frac{1}{4}\mathrm{D}((d\mathbf{y})'\boldsymbol{A} d\mathbf{y}) + \mathrm{cov}(\boldsymbol{a}'(d\mathbf{y}), (d\mathbf{y})'\boldsymbol{A} d\mathbf{y}), \end{aligned}$$

where $\boldsymbol{a} = \frac{\partial f(\mathbf{y})}{\partial \mathbf{y}'} \,|_{\mathbf{y}=\boldsymbol{\mu}}$, $\boldsymbol{A} = \frac{\partial f^2(\mathbf{y})}{\partial \mathbf{y} \partial \mathbf{y}'} \,|_{\mathbf{y}=\boldsymbol{\mu}}$ *and* $d\mathbf{y} = \triangle\mathbf{y} = \mathbf{y} - \boldsymbol{\mu}$.

In this work this approach will be used in order to obtain approximated variances of estimators of NLSS functions of distribution parameters.

In the following lemma an approximation of the estimators $f(\hat{G}(\boldsymbol{X}))$ and $\hat{g}(f(\boldsymbol{x}_i), i = 1, \cdots, n)$, on the basis of a second degree Taylor series expansion, is presented.

Lemma 2.1

Let $\boldsymbol{\theta}$ represent an unknown parameter vector. Furthermore, $\boldsymbol{\theta}$ can be estimated by functions $\hat{G}$ and $\hat{g}$ in two different ways, as explained in Definition 8.

Now, suppose a NLSS function $f(\boldsymbol{\theta})$ has to be estimated.

An approximation of the estimators $f(\hat{G}(\boldsymbol{X}))$ and $\hat{g}(f(\boldsymbol{x}_i), i = 1, \cdots, n)$, on the basis of a second degree Taylor series expansion, is given by:

$$T_1^{Taylor} = f(\hat{G}(\boldsymbol{X}))^{Taylor} = f(\boldsymbol{\theta}) + [f^{(1)}(\boldsymbol{\theta})]'(\hat{G}(\boldsymbol{X}) - \boldsymbol{\theta}) + \frac{1}{2}(\hat{G}(\boldsymbol{X}) - \boldsymbol{\theta})' f^{(2)}(\boldsymbol{\theta})(\hat{G}(\boldsymbol{X}) - \boldsymbol{\theta})$$

$$T_2^{Taylor} = \hat{g}(f(\boldsymbol{x}_i), i = 1, \cdots, n)^{Taylor} = f(\boldsymbol{\theta}) + [f^{(1)}(\boldsymbol{\theta})]'\hat{g}((\boldsymbol{x}_i - \boldsymbol{\theta}), i = 1, \cdots, n) + \frac{1}{2}\hat{g}((\boldsymbol{x}_i - \boldsymbol{\theta})' f^{(2)}(\boldsymbol{\theta})(\boldsymbol{x}_i - \boldsymbol{\theta}), i = 1, \cdots, n),$$

where $f^{(1)}(\boldsymbol{\theta})$ represents the gradient vector and $f^{(2)}(\boldsymbol{\theta})$ the Hessian matrix of the function $f(\boldsymbol{\theta})$, see Definition 4.

It can be seen that both estimation approaches differ just on the second order polynomial of the approximation.

In the same way, as in Troschke (2002), an approximation on the basis of a linear plus quadratic function, of the estimators T_1 and T_2 is presented in the following lemma.

Lemma 2.2

An approximation of the estimators $f(\hat{G}(\boldsymbol{X}))$ and $\hat{g}(f(\boldsymbol{x}_i), i = 1, \cdots, n)$, on the basis of a linear plus quadratic function, is given by:

$$T_1^{Poly} = f(\hat{G}(\boldsymbol{X}))^{Poly} = a_0 + \boldsymbol{a}'\hat{G}(\boldsymbol{X}) + [\hat{G}(\boldsymbol{X})]'\boldsymbol{A}\hat{G}(\boldsymbol{X})$$

$$T_2^{Poly} = \hat{g}(f(\boldsymbol{x}_i), i = 1, \cdots, n)^{Poly} = a_0 + \boldsymbol{a}'\hat{g}(\boldsymbol{x}_i, i = 1, \cdots, n) + \hat{g}(\boldsymbol{x}_i'\boldsymbol{A}\boldsymbol{x}_i, i = 1, \cdots, n),$$

where a_0 is a constant, $\boldsymbol{a} = f^{(1)}(\boldsymbol{\theta})$ and $\boldsymbol{A} = \frac{1}{2}f^{(2)}(\boldsymbol{\theta})$.

Troschke (2002) pointed out that similar to a higher order Taylor approximation, an approximation on the basis of a linear plus quadratic function may result in more accurate combined prediction.

2.3 General Procedure

This work intends to answer the following questions:

1. Which approach should be preferred, $f(\hat{G}(\boldsymbol{X}))$ or $\hat{g}(f(\boldsymbol{x}_i), i = 1, \cdots, n)$, as an estimator of a NLSS function of a p–dimensional parameter vector, $f(\boldsymbol{\theta})$?
2. Is an improvement of these estimators possible?

To answer these questions the following approach is developed in the next sections:

Suppose a NLSS (of at least class C^2) function $f(\boldsymbol{\theta})$ of parameter(s) $\boldsymbol{\theta}$ of a distribution has to be estimated, then:

1. the function $f(\boldsymbol{\theta})$, will be estimated by means of the estimators $T_1 = f(\hat{G}(\boldsymbol{X}))$ and $T_2 = \hat{g}(f(\boldsymbol{x}_i), i = 1, \cdots, n)$,
2. the generalised Jackknife approach will be used in Section A.1.3 in order to generate an unbiased estimator for a second order Taylor approximation of the NLSS function $f(\boldsymbol{\theta})$,
3. if the unbiased estimator is obtained, a linear adjustment as presented in Troschke (2002), is applied in order to obtain a minimal Mean Squared Error (MSE) estimator. This estimator may be biased, but has smaller MSE than the unbiased one generated by the generalised Jackknife approach.
4. improved approximated expressions for the mean and variance of all the aforementioned estimators will be obtained by using the approach presented in Neudecker and Trenkler (2005a).

Chapter 3

Estimation of NLSS Functions of Distribution Parameters

Contents

In the preceding section a new approach for the estimation of NLSS functions has been presented. In this section, special emphasis will be placed on the estimators $f(\hat{G}(\boldsymbol{X}))$ and $\hat{g}(f(\boldsymbol{x}_i), i = 1, \cdots, n)$ for NLSS functions of parameters of multivariate and univariate distributions. For both the multivariate and univariate distributions, the normal distribution will also be considered, and for the multivariate distributions the properties of estimators for correlated random variables will be considered as a special case. Approximated means and variances of the aforementioned estimators, as well as an approximation of the covariance between them, are presented in this section.

3.1 Estimation of NLSS Functions of Parameters of Multivariate Distributions

In multivariate statistics, estimation of NLSS functions of parameters, such as the vector of means $\boldsymbol{\mu}$ of a particular distribution occurs commonly. Therefore, starting from here this work will focus on the estimation of NLSS functions of this parameter, $f(\boldsymbol{\mu})$.

For the estimation of $f(\boldsymbol{\mu})$, under simple random schemes, the usual choices are to calculate first the arithmetic mean and then apply the function afterwards, i.e. compute $f(\hat{G}(\boldsymbol{X})) = f(\bar{\boldsymbol{x}})$; or to do it the other way round, $\hat{g}(f(\boldsymbol{x}_i), i = 1, \ldots, n) = \frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{x}_i)$, where $\boldsymbol{x}_i$, $i = 1, \ldots, n$ denotes the i-th row of the sample matrix $\boldsymbol{X}$, see Definition 8. Moreover, $\bar{\boldsymbol{x}}$ is given in Definition 7.

As in Lemma 2.1 the functions $f(\bar{\boldsymbol{x}})$ and $\frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{x}_i)$ can be approximated by:

$$T_1^{Taylor} = f^{Taylor}(\bar{\boldsymbol{x}}) = f(\boldsymbol{\mu}) + [f^{(1)}(\boldsymbol{\mu})]'(\bar{\boldsymbol{x}} - \boldsymbol{\mu}) + \frac{1}{2}(\bar{\boldsymbol{x}} - \boldsymbol{\mu})' f^{(2)}(\boldsymbol{\mu})(\bar{\boldsymbol{x}} - \boldsymbol{\mu})$$

$$T_2^{Taylor} = \frac{1}{n}\sum_{i=1}^{n} f^{Taylor}(\boldsymbol{x}_i) = f(\boldsymbol{\mu}) + [f^{(1)}(\boldsymbol{\mu})]'\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu}) + \frac{1}{2n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})' f^{(2)}(\boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu}),$$

where $f^{(1)}(\boldsymbol{\mu})$ represents the gradient vector and $f^{(2)}(\boldsymbol{\mu})$ the Hessian matrix of the function $f(\boldsymbol{\mu})$, as presented in Definition 4.

For the particular case of the estimation of $f(\boldsymbol{\mu})$, an approximation by linear plus quadratic forms of the function itself as well as of the estimators $f(\bar{\boldsymbol{x}})$ and $\frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{x}_i)$ is presented in the following remark.

Remark 3.1

Let a_0 be a constant, $\boldsymbol{a}$ a non-stochastic vector and $\boldsymbol{A}$ a non-stochastic symmetric matrix. Let $\boldsymbol{\mu}$ be the vector of means of a p-dimensional distribution, then a NLSS function of this parameter vector, $f(\boldsymbol{\mu})$, can be approximated by:

$$f^{Poly}(\boldsymbol{\mu}) = a_0 + \boldsymbol{a}'\boldsymbol{\mu} + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}.$$

In the same way, the functions $f(\bar{\boldsymbol{x}})$ *and* $\frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{x}_i)$ *can be approximated by the following linear plus quadratic functions in* $\bar{\boldsymbol{x}}$ *and* $\boldsymbol{x}_i$*, respectively:*

$$\begin{aligned} T_1^{Poly} = \; f^{Poly}(\bar{\boldsymbol{x}}) &= a_0 + \tfrac{1}{n}\boldsymbol{a}'\boldsymbol{X}'\mathbf{1}_n + \tfrac{1}{n^2}\mathbf{1}_n'\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}'\mathbf{1}_n \\ &= a_0 + \tfrac{1}{n}\boldsymbol{a}'\boldsymbol{X}'\mathbf{1}_n + \tfrac{1}{n^2}tr\mathbf{1}_n\mathbf{1}_n'\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}' \\ &= a_0 + \boldsymbol{a}'\bar{\boldsymbol{x}} + \bar{\boldsymbol{x}}'\boldsymbol{A}\bar{\boldsymbol{x}} \end{aligned}$$

and

$$T_2^{Poly} = \frac{1}{n}\sum_{i=1}^{n} f^{Poly}(\boldsymbol{x}_i) = a_0 + \boldsymbol{a}'\bar{\boldsymbol{x}} + \tfrac{1}{n}\Sigma_{i=1}^{n}\boldsymbol{x}_i'\boldsymbol{A}\boldsymbol{x}_i,$$

where $\boldsymbol{a} = f^{(1)}(\boldsymbol{\mu})$ *and* $\boldsymbol{A} = \frac{1}{2}f^{(2)}(\boldsymbol{\mu})$.

Frauendorf and Trenkler (1998) have shown that the equality $T_1^{Poly} = T_2^{Poly}$ holds if and only if: $tr\boldsymbol{A}\boldsymbol{X}'\boldsymbol{H}\boldsymbol{X} = 0$, where $\boldsymbol{H} = \boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'$, where $\boldsymbol{I}_n$ represents the Identity Matrix and $\mathbf{1}_n$ the n-vector of ones.

In order to represent the approximations T_1^{Poly} and T_2^{Poly} based on all observations $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$ and not on a particular $\boldsymbol{x}_i$, $i = 1, \ldots, n$ a new notation will be introduced in the following remark.

Remark 3.2

Let $\boldsymbol{y} = \mathrm{Vec}\boldsymbol{X}'$ *be the* Vec *operator of the matrix* $\boldsymbol{X}'$*, i.e. a column vector obtained by stacking the column vectors of* $\boldsymbol{X}' = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$ *below one another. The mean and covariance of* $\mathrm{Vec}\boldsymbol{X}'$ *are* $\mathrm{E}[\mathrm{Vec}\boldsymbol{X}'] = \mathbf{1}_n \otimes \boldsymbol{\mu}$ *and* $\mathrm{cov}(\mathrm{Vec}\boldsymbol{X}') = \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}$.

For $\boldsymbol{y} = \mathrm{Vec}\boldsymbol{X}'$*, the approximations* T_1^{Poly} *and* T_2^{Poly}*, presented in Remark 3.1, can also be written as presented in Magnus and Neudecker (1988, Section 2.4) as follows:*

$$T_1^{Poly} = f_0 + \boldsymbol{f}'\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{F}_1\boldsymbol{y}$$

$$T_2^{Poly} = f_0 + \boldsymbol{f}'\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{F}_2\boldsymbol{y},$$

where f_0 *is a constant,* $\boldsymbol{f} = \frac{1}{n}\mathbf{1}_n \otimes \boldsymbol{a}$ *is a non-stochastic vector,* $\boldsymbol{F}_1 = \frac{1}{n^2}\mathbf{1}_n\mathbf{1}_n' \otimes \boldsymbol{A}$ *and* $\boldsymbol{F}_2 = \frac{1}{n}\boldsymbol{I}_n \otimes \boldsymbol{A}$ *are non-stochastic symmetric matrices. Furthermore,* $\otimes$ *represents the Kronecker product.*

For properties of the Vec operator and the Kronecker product, see Appendix A.4.2 and A.4.1, respectively and for relations between them see Appendix A.4.2.

The aforementioned non-stochastic quantities can be represented as:

$f_0 = a_0$, $\boldsymbol{f} = \frac{1}{n}(\boldsymbol{a}' \cdots \boldsymbol{a}')'$,

$$\boldsymbol{F}_1 = \frac{1}{n^2}\begin{pmatrix} \boldsymbol{A} & \boldsymbol{A} & \dots & \boldsymbol{A} & \boldsymbol{A} \\ \boldsymbol{A} & \boldsymbol{A} & \dots & \boldsymbol{A} & \boldsymbol{A} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{A} & \boldsymbol{A} & \dots & \boldsymbol{A} & \boldsymbol{A} \\ \boldsymbol{A} & \boldsymbol{A} & \dots & \boldsymbol{A} & \boldsymbol{A} \end{pmatrix} \quad \text{and} \quad \boldsymbol{F}_2 = \frac{1}{n}\begin{pmatrix} \boldsymbol{A} & \boldsymbol{0} & \dots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{0} & \boldsymbol{A} \end{pmatrix},$$

where a_0 is a constant, $\boldsymbol{a} = f^{(1)}(\boldsymbol{\mu})$ and $\boldsymbol{A} = \frac{1}{2}f^{(2)}(\boldsymbol{\mu})$.

However, when calculating $\text{var}(T_1^{Poly})$ and $\text{var}(T_2^{Poly})$ (the approximated variances of estimators T_1 and T_2, respectively), a relationship between the 3rd and 4th moment matrices of the $\boldsymbol{x}_i$ and $\text{Vec}\boldsymbol{X}'$ has to be established, see Definitions 26 and 27.

This relationship is presented in Lemma A.1.

The matrices $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ as well as $\boldsymbol{\Phi}_*$ and $\boldsymbol{\Psi}_*$ are presented in more detail in Appendix A.5.2. These matrices are useful for the calculation of covariances as well as variances of linear plus quadratic functions.

In Remark A.4 and Lemma A.2 different identities introduced by Seber (1977) and Kleffe and Rao (1988, Section 2.1) are presented. They are useful for the computation of means and variances of linear plus quadratic functions as well as the covariance between two linear plus quadratic functions, Definition 5.

Notice that the expressions presented in Remark A.4 refer to $\boldsymbol{x}_i = \boldsymbol{z}_i + \boldsymbol{\mu}$, i.e. to an arbitrary element of the i.i.d random sample $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n$. In Lemma A.2 the expressions corresponding to the whole sample are presented. For this the expression $\boldsymbol{y} = \text{Vec}\boldsymbol{X}'$ as given in Remark 3.2 is useful.

Neudecker and Trenkler (2002) obtained the following expressions for the expected values and variances of T_1^{Poly} and T_2^{Poly}, when X_1 and X_2 are **uncorrelated**:

(i) $\mathrm{E}[T_1^{Poly}] = a_0 + \boldsymbol{a}'\boldsymbol{\mu} + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu} + \frac{1}{n}tr\boldsymbol{A}\boldsymbol{\Sigma}$

(ii) $\mathrm{E}[T_2^{Poly}] = a_0 + \boldsymbol{a}'\boldsymbol{\mu} + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu} + tr\boldsymbol{A}\boldsymbol{\Sigma}$

(iii) $\mathrm{var}(T_1^{Poly}) = \frac{1}{n^2}[4tr(\boldsymbol{\mu}'\boldsymbol{A}\otimes\boldsymbol{A})\boldsymbol{\Phi} + 2tr(\boldsymbol{a}'\otimes\boldsymbol{A})\boldsymbol{\Phi} + \frac{1}{n}tr(\boldsymbol{A}\otimes\boldsymbol{A})\boldsymbol{\Psi} + 2\frac{n-1}{n}\beta - \frac{\alpha}{n}] + \gamma$

(iv) $\mathrm{var}(T_2^{Poly}) = \frac{1}{n}[4tr(\boldsymbol{\mu}'\boldsymbol{A}\otimes\boldsymbol{A})\boldsymbol{\Phi} + 2tr(\boldsymbol{a}'\otimes\boldsymbol{A})\boldsymbol{\Phi} + tr(\boldsymbol{A}\otimes\boldsymbol{A})\boldsymbol{\Psi} - \alpha] + \gamma$,

with $\gamma = \frac{1}{n}(2\boldsymbol{A}\boldsymbol{\mu}+\boldsymbol{a})'\boldsymbol{\Sigma}(2\boldsymbol{A}\boldsymbol{\mu}+\boldsymbol{a})$, $\alpha = (tr\boldsymbol{A}\boldsymbol{\Sigma})^2$ and $\beta = tr(\boldsymbol{A}\boldsymbol{\Sigma})^2$.

From expressions (i) and (ii) it can be seen that both, T_1^{Poly} and T_2^{Poly} are biased for $f^{Poly}(\boldsymbol{\mu})$, see Remark 3.1. Their bias terms are given by $\mathrm{b}(T_1^{Poly}, f^{Poly}(\boldsymbol{\mu})) = \frac{1}{n}tr\boldsymbol{A}\boldsymbol{\Sigma}$ and $\mathrm{b}(T_2^{Poly}, f^{Poly}(\boldsymbol{\mu})) = tr\boldsymbol{A}\boldsymbol{\Sigma}$, see Definition 20. Notice that T_1^{Poly} is asymptotically unbiased.

Using the identities of Lemma A.2, equivalent expressions to those introduced in Neudecker and Trenkler (2002) for the variances of T_1^{Poly} and T_2^{Poly} as well as the expression for the covariance between T_1^{Poly} and T_2^{Poly} have been deduced in this work and are presented in the following theorem.

Theorem 3.1

Using the properties presented in Remark A.4 and Lemma A.2, the expression for the covariance between T_1^{Poly} and T_2^{Poly} as well as equivalent expressions for their variances to those introduced in Neudecker and Trenkler (2002) can be rewritten as follows:

(i) $\mathrm{var}(T_1^{Poly}) = \frac{1}{n^2}[4tr\boldsymbol{A}\Phi^*(\boldsymbol{A})\boldsymbol{\mu}' + 2tr\boldsymbol{A}\Phi(\boldsymbol{a}) + \frac{1}{n}tr\boldsymbol{A}\Psi(\boldsymbol{A}) + 2\frac{n-1}{n}\beta - \frac{\alpha}{n}] + \gamma$

(ii) $\mathrm{var}(T_2^{Poly}) = \frac{1}{n}[4tr\boldsymbol{A}\Phi^*(\boldsymbol{A})\boldsymbol{\mu}' + 2tr\boldsymbol{A}\Phi(\boldsymbol{a}) + tr\boldsymbol{A}\Psi(\boldsymbol{A}) - \alpha] + \gamma$

(iii) $\mathrm{cov}(T_1^{Poly}, T_2^{Poly}) = \frac{1}{n^2}[2(n+1)tr\boldsymbol{A}\Phi^*(\boldsymbol{A})\boldsymbol{\mu}' + (n+1)tr\boldsymbol{A}\Phi(\boldsymbol{a}) + tr\boldsymbol{A}\Psi(\boldsymbol{A}) - \alpha] + \gamma.$

Proof: See Appendix C.2.

Expressions of the MSE of the approximations T_1^{Poly} and T_2^{Poly} are obtained by using Definition 22, i.e.:

$\mathrm{MSE}(T_l^{Poly}, f^{Poly}(\boldsymbol{\mu})) = \mathrm{E}[(T_l^{Poly} - f^{Poly}(\boldsymbol{\mu}))^2] = \mathrm{var}(T_l^{Poly}) + [\mathrm{b}(T_l^{Poly}, f^{Poly}(\boldsymbol{\mu}))]^2$, $l = 1, 2$.

3.1.1 Properties of Estimators for Correlated Random Variables

Existing methods for generating correlated random variables do not allow the setting of the correlation level between those variables. Therefore, it is necessary to develop a new approach for generating correlated random variables with user-defined correlation level (ρ) and probability distribution. This approach will be presented in Section 6.1.1.

The expressions given in Remark A.5 are based on an arbitrary element of the correlated sample $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$. Now, based on these expressions those corresponding to the whole sample have been deduced in this work and are presented Lemma A.3. These expressions will make it possible to compare the MSEs of the approximations T_1^{Poly} and T_2^{Poly} for correlated random variables.

Neudecker and Trenkler (2002) calculated the variances of the approximations, T_1^{Poly} and T_2^{Poly}, for **uncorrelated** variables, X_1 and X_2. For the calculation of $\text{var}(T_1^{Poly})$ and $\text{var}(T_2^{Poly})$, when the involved variables are **correlated**, a new approach based on Kleffe and Rao (1988, Section 2.1), see Lemma A.3, is used. The expressions for the variances of T_1^{Poly} and T_2^{Poly} for **correlated** random variables, X_1 and X_2 are presented as follows.

Theorem 3.2

Based on Lemma A.3 and the identities exposed in Remark A.5, the following expression for the covariance between T_1^{Poly} and T_2^{Poly} as well as the expressions for their variances, for correlated variables, were obtained in this work:

(i) $\text{var}(T_1^{Poly}) = \frac{1}{n^2}[4tr\boldsymbol{A}\Phi_u^*(\boldsymbol{A})\boldsymbol{\mu}' + 2tr\boldsymbol{A}\Phi_u(\boldsymbol{a}) + tr\boldsymbol{A}\Psi_u(\boldsymbol{A}) + 2\frac{n-1}{n}\beta - \frac{\alpha}{n}] + \gamma$

(ii) $\text{var}(T_2^{Poly}) = \frac{1}{n}[4tr\boldsymbol{A}\Phi_u^*(\boldsymbol{A})\boldsymbol{\mu}' + 2tr\boldsymbol{A}\Phi_u(\boldsymbol{a}) + tr\boldsymbol{A}\Psi_u(\boldsymbol{A}) - \alpha] + \gamma$

(iii) $\text{cov}(T_1^{Poly}, T_2^{Poly}) = \frac{1}{n^2}[2(n+1)tr\boldsymbol{A}\Phi_u^*(\boldsymbol{A})\boldsymbol{\mu}' + (n+1)tr\boldsymbol{A}\Phi_u(\boldsymbol{a}) + tr\boldsymbol{A}\Psi_u(\boldsymbol{A}) - \alpha] + \gamma.$

Proof: See Appendix C.4.

3.1.2 Estimation of NLSS Functions of Parameters of Multinormal Distributions

Now it is assumed that $\boldsymbol{y} = \mathrm{Vec}\boldsymbol{X}'$ is multinormally distributed with $\mathrm{E}[\boldsymbol{y}] = \boldsymbol{g}_* = \mathbf{1}_n \otimes \boldsymbol{\mu}$ and $\mathrm{cov}(\boldsymbol{y}) = \boldsymbol{V}_* = \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}$.

Remark 3.3

For the normal case $\boldsymbol{z}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ are the odd moments equal zero, i.e. $\boldsymbol{\Phi} = \boldsymbol{\Phi}^N = \mathbf{0}$. Furthermore, in Magnus and Neudecker (1979, Theorem 4.3, (iv)) the following expression for the fourth moment is given:

$$\boldsymbol{\Psi} = \boldsymbol{\Psi}^N = (\boldsymbol{I}_{p^2} + \boldsymbol{K}_{pp})(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) + (\mathrm{Vec}\boldsymbol{\Sigma})(\mathrm{Vec}\boldsymbol{\Sigma})'.$$

The index N stands for representing the normal distribution.

Matrices $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$, $\boldsymbol{I}_{p^2} = \boldsymbol{I}_p \otimes \boldsymbol{I}_p$ and $\boldsymbol{K}_{pp}$ have been introduced in Lemma A.1 and are presented in more details in Appendix A.4.

Given $\boldsymbol{y} = \mathrm{Vec}\ \boldsymbol{X}' \sim N(\mathbf{1}_n \otimes \boldsymbol{\mu}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$ and the following identities:

$$\boldsymbol{F}_1\boldsymbol{V}_*\boldsymbol{F}_2\boldsymbol{V}_* = \tfrac{1}{n^3}\mathbf{1}_n\mathbf{1}_n^{'} \otimes (\boldsymbol{A}\boldsymbol{\Sigma})^2$$

$$tr\boldsymbol{F}_1\boldsymbol{V}_*\boldsymbol{F}_2\boldsymbol{V}_* = \tfrac{1}{n^2}tr(\boldsymbol{A}\boldsymbol{\Sigma})^2$$

$$\boldsymbol{F}_l\boldsymbol{g}_* = \tfrac{1}{n}\mathbf{1}_n \otimes \boldsymbol{A}\boldsymbol{\mu},\ l = 1, 2$$

$$(2\boldsymbol{F}_1\boldsymbol{g}_* + \boldsymbol{f})'\boldsymbol{V}_*(2\boldsymbol{F}_2\boldsymbol{g}_* + \boldsymbol{f}) = \tfrac{1}{n}(2\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{a})'\boldsymbol{\Sigma}(2\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{a}) = \gamma,$$

then the following expressions for the variances of T_1^{Poly} and T_2^{Poly} as well as covariance between T_1^{Poly} and T_2^{Poly} were obtained:

(i) $\mathrm{var}(T_1^{Poly}) = \frac{2}{n^2}\beta + \gamma$

(ii) $\mathrm{var}(T_2^{Poly}) = \frac{2}{n}\beta + \gamma$

(iii) $\mathrm{cov}(T_1^{Poly}, T_2^{Poly}) = \mathrm{var}(T_1^{Poly}) = \frac{2}{n^2}\beta + \gamma.$

With $\gamma = \frac{1}{n}(2\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{a})'\boldsymbol{\Sigma}(2\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{a})$, $\quad \alpha = (tr\boldsymbol{A}\boldsymbol{\Sigma})^2$ and $\beta = tr(\boldsymbol{A}\boldsymbol{\Sigma})^2$.

3.2 Estimation of NLSS Functions of Parameters of Univariate Distributions

For the univariate case it is supposed that $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)'$ represents a realisation of an i.i.d. random sample drawn from a probability distribution with mean μ and variance σ^2. In this section the interest is also concentrated on the estimation of the NLSS function $f(\mu)$, where μ represents the mean of a given univariate distribution.

The first step of the estimation approach developed in this work and presented in Section 2.3 consists of the approximation of the function to be estimated. As approximation approach the linear plus quadratic form will be used, see Remark 3.1. This approximation is given by:

$$f^{Poly}(\mu) = a_0 + a\mu + A\mu^2,$$

where a_0, a and A are real constants. Furthermore, $a = f^{(1)}(\mu)$ and $A = \frac{1}{2}f^{(2)}(\mu)$, i.e. the first and second derivative of the function $f(\mu)$.

In the same way the estimators $f(\hat{G}(\boldsymbol{x}))$ and $\hat{g}(f(x_i), i = 1, \cdots, n)$ can be approximated by the linear plus quadratic forms, as made in Lemma 2.2. The resulting approximations are denoted by T_1^{Poly} and T_2^{Poly} and presented as follows:

$$T_1^{Poly} = f^{Poly}(\bar{x}) = a_0 + a\bar{x} + A\bar{x}^2, \text{ with } \bar{x} = \frac{1}{n}\mathbf{1}_n'\boldsymbol{x}$$
$$T_2^{Poly} = \frac{1}{n}\sum_{i=1}^{n} f^{Poly}(x_i) = a_0 + a\bar{x} + \frac{A}{n}\sum_{i=1}^{n} x_i^2.$$

Both T_1^{Poly} and T_2^{Poly} are biased for $f^{Poly}(\mu)$, see Theorem 3.4.

The aforementioned approximations are presented in matrix notation as follows:

$$T_1^{Poly} = f_0 + \boldsymbol{f}'\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{F}_1\boldsymbol{x}$$

$$T_2^{Poly} = f_0 + \boldsymbol{f}'\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{F}_2\boldsymbol{x},$$

with $\boldsymbol{x}$ is a realisation of an i.i.d. random sample, f_0 a real constant, $\boldsymbol{f}$ a $n \times 1$ vector and $\boldsymbol{F}_l$, $l = 1, 2$ a $n \times n$ matrix. They are given by: $\boldsymbol{f} = \frac{a}{n}\mathbf{1}_n$, $\boldsymbol{F}_1 = \frac{A}{n^2}\mathbf{1}_n\mathbf{1}_n'$ and $\boldsymbol{F}_2 = \frac{A}{n}\boldsymbol{I}_n$.

Like in the multivariate case, when calculating var(T_1^{Poly}) and var(T_2^{Poly}) a relationship between the 3rd and 4th moment matrices of the x_i and Vec$\boldsymbol{x}'$ has to be established. The third and fourth moment matrices for the univariate case $\boldsymbol{\Phi}_* = E[\boldsymbol{z} \otimes \boldsymbol{z}\boldsymbol{z}']$ and $\boldsymbol{\Psi}_* = E[\boldsymbol{z}\boldsymbol{z}' \otimes \boldsymbol{z}\boldsymbol{z}']$ are given in the following theorem.

Theorem 3.3

Assume that $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)'$ represents a realisation of an i.i.d. random sample drawn from a probability distribution with mean μ and variance σ^2. Then $\boldsymbol{z} = \boldsymbol{x} - \mathrm{E}[\boldsymbol{x}]$, with $\mathrm{E}[\boldsymbol{z}] = \boldsymbol{0}$ and $\mathrm{E}[\boldsymbol{z}\boldsymbol{z}'] = \mathrm{cov}(\boldsymbol{x}) = \sigma^2 \boldsymbol{I}_n$. For the univariate case $\boldsymbol{\Phi}_*$ and $\boldsymbol{\Psi}_*$ are given by:

- $\boldsymbol{\Phi}_* = (\boldsymbol{I}_n \otimes \boldsymbol{I}_n)\boldsymbol{G}\Phi$, *where* $\boldsymbol{G} = (\boldsymbol{E}_{11}, \ldots, \boldsymbol{E}_{nn})'$, *with* $\boldsymbol{E}_{ii} = \boldsymbol{e}_i\boldsymbol{e}_i'$, $\boldsymbol{e}_i$ *is the i-th member of the canonical basis of* $\mathbb{R}^n$, *and it is known that the commutation matrix* $\boldsymbol{K}_{1,n} = \boldsymbol{K}_n = \boldsymbol{I}_n$, *see Appendix A.4.3.*
- $\boldsymbol{\Psi}_* = (\boldsymbol{I}_{n^2} + \boldsymbol{K}_{nn})(\sigma^4 \boldsymbol{I}_{n^2}) + \sigma^4[\mathrm{Vec}(\boldsymbol{I}_n)][\mathrm{Vec}(\boldsymbol{I}_n)]' + \boldsymbol{I}_{n^2}[\boldsymbol{K}_{nn}(\Psi - 3\sigma^4)]\boldsymbol{I}_{n^2}$, *with* $\boldsymbol{K}_{nn} = \sum_{i=1}^{n}(\boldsymbol{E}_{ii} \otimes \boldsymbol{E}_{ii})$ *and* $\boldsymbol{I}_{n^2} = (\boldsymbol{I}_n \otimes \boldsymbol{I}_n)$.

Proof: Follows from Neudecker and Trenkler (2002), Theorem 1, where the same properties for the multivariate case are presented.

According to the assumption on the rows of the vector $\boldsymbol{x}$, it has already been mentioned that the z_i, $i = 1, \ldots, n$, occurring in the vector $\boldsymbol{z} = (z_1, \ldots z_n)'$ are mutually uncorrelated, with $\mathrm{E}[z_i] = 0$ and $\mathrm{var}(z_i) = \mathrm{E}[z_i^2] = \sigma^2$, $i = 1, \ldots, n$.

For the univariate case the variance of T_1^{Poly} and T_2^{Poly} as well as cov(T_1^{Poly}, T_2^{Poly}) can be obtained using the expressions presented in Neudecker and Trenkler (2002) or those presented in Kleffe and Rao (1988). The identities corresponding to the approach given by Neudecker and Trenkler (2002) are presented in Appendix C.5.

For the obtention of the variances of T_l^{Poly}, $l = 1, 2$ as well as cov(T_1^{Poly}, T_2^{Poly}) following the approach presented in Kleffe and Rao (1988) the identities summarised in Lemma A.4 will be useful. The expressions for the means and the variances of T_l^{Poly} for $l = 1, 2$ and the covariance between T_1^{Poly} and T_2^{Poly} are presented in the following theorem.

Theorem 3.4 (Mean and Variance of T_1^{Poly} and T_2^{Poly} for Univariate Distributions)

Let $\Phi = \mathrm{E}[z_i^3] = \mathrm{E}[x_i - \mu]^3$ and $\Psi = \mathrm{E}[z_i^4] = \mathrm{E}[x_i - \mu]^4$ represent the third and fourth central moments of x_i, respectively. The expressions for the means and the variances of T_1^{Poly} and T_2^{Poly} and for $\mathrm{cov}(T_1^{Poly}, T_2^{Poly})$ for the univariate case are the following:

(i) $\mathrm{E}(T_1^{Poly}) = a_0 + a\mu + A\mu^2 + \frac{1}{n}A\sigma^2$

(ii) $\mathrm{E}(T_2^{Poly}) = a_0 + a\mu + A\mu^2 + A\sigma^2$

(iii) $\mathrm{var}(T_1^{Poly}) = \frac{1}{n^2}[4A^2\mu\Phi + 2aA\Phi + \frac{1}{n}A^2\Psi + \frac{1}{n}(2n-3)\alpha] + \gamma$

(iv) $\mathrm{var}(T_2^{Poly}) = \frac{1}{n}[4A^2\mu\Phi + 2aA\Phi + A^2\Psi - \alpha] + \gamma$

(v) $\mathrm{cov}(T_1^{Poly}, T_2^{Poly}) = \frac{1}{n^2}[2(n+1)A^2\mu\Phi + (n+1)aA\Phi + A^2\Psi - \alpha] + \gamma,$

with common terms $\gamma = \frac{\sigma^2}{n}(2A\mu + a)^2$ and $\alpha = \beta = (A\sigma^2)^2 = A^2\sigma^4$.

Proof: See Appendix C.6.

Notice that $[\mathrm{b}(T_1^{Poly}, f^{Poly(\mu)}]^2 = \frac{\alpha}{n^2}$ and $[\mathrm{b}(T_2^{Poly}, f^{Poly(\mu)}]^2 = \alpha$, the bias of the approximations T_1^{Poly} and T_2^{Poly}, respectively.

Theorem 3.5 (Estimation of NLSS Functions for Normal Distributions)

Assume that $\boldsymbol{x}$ is normally distributed with $\mathrm{E}[\boldsymbol{x}] = \boldsymbol{g}_* = \mu\mathbf{1}_n$ and $\mathrm{cov}(\boldsymbol{x}) = \boldsymbol{V}_* = \sigma^2\boldsymbol{I}_n$. $\boldsymbol{x} \sim N(\mu\mathbf{1}_n, \sigma^2\boldsymbol{I}_n)$, with $\Phi = \Phi^N = 0$ and $\Psi = \Psi^N = 3\sigma^2$. The index N stands for the normal distribution.

Using the expressions given in the multivariate case and using the following identities:

$$\boldsymbol{F}_1\boldsymbol{V}_*\boldsymbol{F}_2\boldsymbol{V}_* = \frac{A^2\sigma^4}{n^3}\mathbf{1}_n\mathbf{1}_n^{'}$$

$$tr\boldsymbol{F}_1\boldsymbol{V}_*\boldsymbol{F}_2\boldsymbol{V}_* = \frac{\alpha}{n^2}$$

$$\boldsymbol{F}_l\boldsymbol{g}_* = \frac{A\mu}{n}\mathbf{1}_n,\ l = 1, 2$$

$$(2\boldsymbol{F}_l\boldsymbol{g}_* + \boldsymbol{f})'\boldsymbol{V}_*(2\boldsymbol{F}_2\boldsymbol{g}_* + \boldsymbol{f}) = \frac{\sigma^2}{n}(2A\mu + a)^2 = \gamma,$$

the following expressions for the variances of T_1^{Poly} and T_2^{Poly} as well as covariance between T_1^{Poly} and T_2^{Poly} are obtained:

(i) $\mathrm{var}\ (T_1^{Poly}) = \frac{2}{n^2}\beta + \gamma$

(ii) $\mathrm{var}\ (T_2^{Poly}) = \frac{2}{n}\beta + \gamma$

(iii) $\mathrm{cov}(T_1^{Poly}, T_2^{Poly}) = \mathrm{var}(T_1^{Poly}) = \frac{2}{n^2}\beta + \gamma$, *with $\beta = A^2\sigma^4$.*

Chapter 4

Unbiased Estimation of NLSS Functions of Distribution Parameters

Contents

As mentioned in Section 2.3 the new inference method developed in this work is based on the construction of an unbiased estimator for the approximation of $f(\boldsymbol{\theta})$ using the generalised Jackknife approach. This approach consists of a linear combination of the approximation of two existing estimators, in this case T_1^{Poly} and T_2^{Poly}, see Appendix A.1.3.

In Section 3, the properties T_1^{Poly} and T_2^{Poly} have been presented. In this section, the mathematical foundations of the new estimation approach as well as its properties are presented.

4.1 Unbiased Estimation of NLSS Functions of Parameters of Multivariate Distributions

Let $\boldsymbol{X}$ denote the sample matrix, i.e. $\boldsymbol{X}' = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$, where $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ represents a realisation of an i.i.d. random sample drawn from a p–dimensional probability distribution with mean $\boldsymbol{\mu} = E[\boldsymbol{x}_i]$ and covariance matrix $\boldsymbol{\Sigma} = E[(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})']$.

In Section 3, approximations of estimators $T_1 = f(\hat{G}(\boldsymbol{X}))$ and $T_2 = \hat{g}(F(\boldsymbol{X}))$ were presented and their properties were investigated.

Using the Jackknife procedure described in Section A.1.3, an unbiased estimator for the function $f(\boldsymbol{\mu})$ can be obtained.

This estimator is constructed as follows:

$$T_3 = \frac{T_1 - RT_2}{1 - R} = \frac{1}{n-1}(n\, T_1 - T_2), \quad \text{for} \quad R = \frac{\mathrm{b}(T_1, f(\boldsymbol{\mu}))}{\mathrm{b}(T_2, f(\boldsymbol{\mu}))} = \frac{1}{n}, \tag{4.1.1}$$

where $\mathrm{b}(T_l, f(\boldsymbol{\mu}))$, $l = 1, 2$ represent the bias terms of the estimators T_1 and T_2.

Unfortunately, exact expressions for $\mathrm{b}(T_l, f(\boldsymbol{\mu}))$, $l = 1, 2$ are not known.

In this case, the approximations, on the basis of a linear plus quadratic function, of the estimators T_1 and T_2 and their properties will be used in order to construct an unbiased estimator for the approximation $f^{Poly}(\boldsymbol{\mu})$.

In Section 2.2.3 it was shown that T_1^{Poly} and T_2^{Poly} are biased for $f^{Poly}(\boldsymbol{\mu})$. Their expected values have been presented in Section 3.1.

Their bias with respect to $f^{Poly}(\boldsymbol{\mu})$ are given by:

$\mathrm{b}(T_1^{Poly}, f^{Poly}(\boldsymbol{\mu})) = \frac{1}{n} tr\boldsymbol{A\Sigma}$, and

$\mathrm{b}(T_2^{Poly}, f^{Poly}(\boldsymbol{\mu})) = tr\boldsymbol{A\Sigma}$, respectively.

Using the Jackknife procedure described in Section A.1.3, Neudecker and Trenkler (2002) introduced an unbiased estimator for the approximation $f^{Poly}(\boldsymbol{\mu})$, denoted by T_3^{Poly}.

This estimator is constructed as follows:

$$T_3^{Poly} = \frac{T_1^{Poly} - RT_2^{Poly}}{1-R} = \frac{1}{n-1}(n\, T_1^{Poly} - T_2^{Poly}), \tag{4.1.2}$$

for $R = \dfrac{\mathrm{b}(T_1^{Poly}, f^{Poly}(\boldsymbol{\mu}))}{\mathrm{b}(T_2^{Poly}, f^{Poly}(\boldsymbol{\mu}))} = \dfrac{1}{n}$.

Based on Lemma 2.2, T_3^{Poly} is given by:

$$T_3^{Poly} = a_0 + \boldsymbol{a}'\bar{\boldsymbol{x}} + \frac{1}{n(n-1)}\sum_{i\neq k}\boldsymbol{x}_i'\boldsymbol{A}\boldsymbol{x}_k. \tag{4.1.3}$$

Remark 4.1

Notice that R does not depend on unknown quantities and is a function of $1/n^m$ with $m = 1$, see Remark A.2. For T_1^{Poly} and T_2^{Poly} given as in Remark 3.2, T_3^{Poly} is unbiased for $f^{Poly}(\boldsymbol{\mu})$.

This is shown as follows:

$$\begin{aligned}
\mathrm{b}(T_3^{Poly}, f^{Poly}(\boldsymbol{\mu})) &= \tfrac{n}{n-1}[\mathrm{b}(T_1^{Poly}, f^{Poly}(\boldsymbol{\mu})) - \mathrm{b}(T_2^{Poly}, f^{Poly}(\boldsymbol{\mu}))] + \mathrm{b}(T_2, f^{Poly}(\boldsymbol{\mu}))\\
&= \tfrac{1}{n-1}[n\mathrm{b}(T_1^{Poly}, f^{Poly}(\boldsymbol{\mu})) - \mathrm{b}(T_2^{Poly}, f^{Poly}(\boldsymbol{\mu}))]\\
&= \tfrac{1}{n-1}[ntr\boldsymbol{A\Sigma}/n - tr\boldsymbol{A\Sigma}] = 0.
\end{aligned}$$

T_3^{Poly} can also be written in matrix notation as in Magnus and Neudecker (1988, Section 2.4), i.e.

$$T_3^{Poly} = f_0 + \boldsymbol{f}'\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{F}_3\boldsymbol{y}, \text{ with } \boldsymbol{F}_3 = -\tfrac{1}{n(n-1)}\boldsymbol{L}\otimes\boldsymbol{A} \text{ and } \boldsymbol{L} = \boldsymbol{I}_n - \boldsymbol{1}_n\boldsymbol{1}_n'.$$

Neudecker and Trenkler (2002) obtained the following expressions for the variance of T_3^{Poly} when the variables involved, i.e. X_1 and X_2, are **uncorrelated**:

$$\mathrm{var}(T_3^{Poly}) = \frac{2}{n(n-1)}\beta + \gamma, \text{ with } \gamma = \frac{1}{n}(2\boldsymbol{A\mu} + \boldsymbol{a})'\boldsymbol{\Sigma}(2\boldsymbol{A\mu}+\boldsymbol{a}) \text{ and } \beta = tr(\boldsymbol{A\Sigma})^2.$$

Since T_3^{Poly} is unbiased for $f^{Poly}(\boldsymbol{\mu})$, i.e. $\mathrm{b}(T_3^{Poly}, f^{Poly}(\boldsymbol{\mu})) = 0$, it holds:

$$\mathrm{MSE}(T_3^{Poly}, f^{Poly}(\boldsymbol{\mu})) = E[(T_3^{Poly} - f^{Poly}(\boldsymbol{\mu}))^2] = \mathrm{var}(T_3^{Poly}).$$

Since $\mathrm{var}(T_3^{Poly})$ does not depend on the third and fourth moments matrices $\boldsymbol{\Phi} = \mathrm{E}[\boldsymbol{z}_i \otimes \boldsymbol{z}_i \boldsymbol{z}_i']$ and $\boldsymbol{\Psi} = \mathrm{E}[\boldsymbol{z}_i \boldsymbol{z}_i' \otimes \boldsymbol{z}_i \boldsymbol{z}_i']$, $i = 1, \cdots, n$, see Lemma A.1, this expression is the same for the multinormal case and for the case where the variables involved are correlated.

4.2 Unbiased Estimation of NLSS Functions of Parameters of Univariate Distributions

For the univariate case it is supposed that $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)'$ represents a realisation of an i.i.d. random sample drawn from a probability distribution with mean μ and variance σ^2.

Based on the generalised Jackknife approach the new estimator is given by:

$$T_3^{Poly} = a_0 + a\bar{x} - \sum_{i \neq j}^{n} x_i A x_j.$$

This estimator can also be written in matricial notation as presented in Magnus and Neudecker (1988, Section 2.4), i.e.

$$T_3^{Poly} = f_0 + \boldsymbol{f}'\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{F}_3\boldsymbol{x},$$

where $\boldsymbol{f} = \frac{a}{n}\mathbf{1}_n$ and $\boldsymbol{F}_3 = -\frac{A}{n(n-1)}\boldsymbol{L}$, with $\boldsymbol{L} = \boldsymbol{I}_n - \mathbf{1}_n\mathbf{1}_n'$.

The expression for the variance of T_3^{Poly} for the univariate case is the following:

$$\mathrm{var}(T_3^{Poly}) = \frac{2}{n(n-1)}\beta + \gamma, \text{ where } \gamma = \frac{\sigma^2}{n}(2A\mu + a)^2, \ \beta = \alpha = (A\sigma^2)^2 = A^2\sigma^4.$$

Since $\mathrm{var}(T_3^{Poly})$ does not depend on the third and fourth moments of x_i, i.e. $\Phi = E[z_i^3] = E[x_i - \mu]^3$ and $\Psi = E[z_i^4] = E[x_i - \mu]^4$, this expression is the same for the normal case.

Chapter 5

Minimal MSE Estimation of NLSS Functions of Distribution Parameters

Contents

Following the approach presented in Section 2.2, in this section it will be investigated if an improvement of the presented estimators T_1, T_2 and T_3 can be made. For this, the *unrestricted linear adjustment* presented by Troschke (2002) will be used. The mathematical foundations of the aforementioned approach are presented in this section.

The most important properties, e.g. the respective expressions for the variance and MSE of the resulting estimator will also be presented in this section.

The multivariate and univariate cases are presented separately and the multinormal and normal distributions are investigated as special cases.

5.1 Minimal MSE Estimation of NLSS Functions of Parameters of Multivariate Distributions

In Troschke (2002, Chapter 1.3) the linear plus quadratic approach for the combination of forecasts is introduced. It is shown how the combination of estimators within the corresponding classes should be chosen in order to minimise the Mean Squared Prediction Error (MSPE) of the combined forecasts. The optimal combination of parameters depends on the first to fourth order moments of the joint distribution of the target variable and its forecasts. In the same work it is also widely explained how a *linear adjustment* is applied to find an optimal Mean Squared Prediction Error (MSPE) of forecast F for the target variable y, MSPE(F, y).

In this section the same approach will be used in order to find an optimal κ or κ_{min}, so that MSE$(T_{\kappa_{min}}, f(\boldsymbol{\mu}))= \min_{\kappa}$ MSE$(T_\kappa, f(\boldsymbol{\mu}))$.

As made for estimators T_1, T_2 and T_3, an approximation, on the basis of a linear plus quadratic function, of the estimator $T_{\kappa_{min}}$ will be used in order to obtain approximated expressions for the MSEs of this estimator, i.e. MSE$(T^{Poly}_{\kappa_{min}}, f^{Poly}(\boldsymbol{\mu}))$.

Lemma 5.1

Assume that the estimator T_3^{Poly} can be written as $T_\kappa^{Poly} = \kappa(T_1^{Poly} - T_2^{Poly}) + T_2^{Poly}$, with $\kappa = \frac{n}{n-1}$, then the MSE of T_κ^{Poly} can be minimised using the unrestricted linear adjustment without constant term *investigated by Troschke (2002).*

The estimator T_κ^{Poly} has the following properties:

$$\mathrm{E}[T_\kappa^{Poly}] = \kappa \mathrm{E}[T_1^{Poly}] + (1-\kappa)\mathrm{E}[T_2^{Poly}]$$

$$\mathrm{b}(T_\kappa^{Poly}, f^{Poly}(\boldsymbol{\mu})) = \kappa(\iota_1 - \iota_2) + \iota_2, \text{ where } \iota_l = \mathrm{b}(T_l^{Poly}, f^{Poly}(\boldsymbol{\mu})),\ l = 1, 2$$

$$\mathrm{var}(T_\kappa^{Poly}) = \kappa^2 \mathrm{var}(T_1^{Poly} - T_2^{Poly}) + \mathrm{var}(T_2^{Poly}) + 2\kappa \mathrm{cov}((T_1^{Poly} - T_2^{Poly}), T_2^{Poly})$$

$$\mathrm{MSE}(T_\kappa^{Poly}, f^{Poly}(\boldsymbol{\mu})) = \mathrm{var}(T_\kappa^{Poly}) + [\mathrm{b}(T_\kappa^{Poly}, f^{Poly}(\boldsymbol{\mu}))]^2.$$

This estimator is unbiased for $\kappa = \frac{n}{n-1}$.

Proof:

Since $T_\kappa^{Poly} = \frac{1}{n-1}(n\, T_1^{Poly} - T_2^{Poly})$ it follows immediately that $T_\kappa^{Poly} = \kappa(T_1^{Poly} - T_2^{Poly}) + T_2^{Poly}$, with $\kappa = \frac{n}{n-1}$.

Using these expressions and different identities of mean and variance, see for example Mood, Graybill and Boes (1974), it can be demonstrated that:

$$\begin{aligned} \mathrm{E}[T_\kappa^{Poly}] &= E[\kappa(T_1^{Poly} - T_2^{Poly}) + T_2^{Poly}] = \kappa E[T_1^{Poly} - T_2^{Poly}] + E[T_2^{Poly}] \\ &= \kappa E[T_1^{Poly}] + (1-\kappa)E[T_2^{Poly}]. \end{aligned}$$

For $R = \frac{1}{n}$ the estimator is unbiased as was demonstrated in Section 4.1.

$$\begin{aligned} \mathrm{var}(T_\kappa^{Poly}) &= \mathrm{var}(\kappa(T_1^{Poly} - T_2^{Poly}) + T_2^{Poly}) \\ &= \kappa^2 \mathrm{var}(T_1^{Poly} - T_2^{Poly}) + \mathrm{var}(T_2^{Poly}) + 2\kappa \mathrm{cov}((T_1^{Poly} - T_2^{Poly}), T_2^{Poly}). \end{aligned}$$

Like in Troschke (2002) the following notation is used:

$\varrho_i = \iota_2 - \iota_1$, where $\iota_l = \mathrm{b}(T_l^{Poly}, f^{Poly}(\boldsymbol{\mu}))$, $l = 1, 2$ the bias of estimator T_l^{Poly},

$\varrho_0 = \iota_2$, then it follows $\varrho_0 \varrho_i = \iota_2(\iota_2 - \iota_1)$,

$\Sigma_{ii} = \mathrm{var}(T_1^{Poly} - T_2^{Poly})$,

$\Sigma_{00} = \mathrm{var}(T_2^{Poly})$ and

$-\Sigma_{i0} = \mathrm{cov}((T_1^{Poly} - T_2^{Poly}), T_2^{Poly}) = \mathrm{cov}(T_1^{Poly}, T_2^{Poly}) - \mathrm{var}(T_2^{Poly})$.

Using the notation presented above the bias, the variance and the MSE of T_κ can be rewritten as:

$$\mathrm{b}(T_\kappa^{Poly}, f^{Poly}(\boldsymbol{\mu})) = \varrho_0 - \kappa \varrho_i, \tag{5.1.1}$$

$$\mathrm{var}(T_\kappa^{Poly}) = \kappa^2 \Sigma_{ii} + \Sigma_{00} - 2\kappa \Sigma_{i0}, \tag{5.1.2}$$

$$\mathrm{MSE}(T_\kappa^{Poly}, f^{Poly}(\boldsymbol{\mu})) = \Sigma_{00} + \varrho_0^2 + \kappa^2 \Sigma_{ii} - 2\kappa \Sigma_{i0} + \kappa^2 \varrho_i^2 - 2\kappa \varrho_0 \varrho_i. \tag{5.1.3}$$

According to Troschke (2002, Equation 1.17), the optimal choice for κ obtained using Lemma 5.1 is given by:

$$\kappa_{min} = \frac{\Sigma_{i0} + \varrho_0 \varrho_i}{\Sigma_{ii} + \varrho_i^2} = \frac{\mathrm{var}(T_2^{Poly}) - \mathrm{cov}(T_1^{Poly}, T_2^{Poly}) + \iota_2(\iota_2 - \iota_1)}{\mathrm{var}(T_1^{Poly} - T_2^{Poly}) + (\iota_2 - \iota_1)^2}. \tag{5.1.4}$$

Theorem 5.1

With the results presented in Theorem 3.2, it is possible to deduce κ_{min} by means of Equation 5.1.4. This quantity is given by:

$$\kappa_{min} = \frac{n[2tr\boldsymbol{A}\Phi^*(\boldsymbol{A})\boldsymbol{\mu}' + tr\boldsymbol{A}\Phi(\boldsymbol{a}) + tr\boldsymbol{A}\Psi(\boldsymbol{A}) + (n-1)\alpha]}{(n-1)tr\boldsymbol{A}\Psi(\boldsymbol{A}) + 2\beta + (n-1)^2\alpha},$$

with $\Phi^*(\boldsymbol{A})$, $\Phi(\boldsymbol{a})$ and $\Psi(\boldsymbol{A})$ as given in Remark A.4.

Proof: See Appendix C.7.

When correlated random variables are analysed κ_{min} is also given by this equation with different expressions for $\mathrm{var}(T_l^{Poly})$, $l = 1, 2$ and $\mathrm{cov}(T_1^{Poly}, T_2^{Poly})$, see Theorem 3.2.

Theorem 5.2

Let κ_{min} be given as in Theorem 5.1, the MSE of $T_{\kappa_{min}}^{Poly}$ is given by:

$$\begin{aligned} \mathrm{MSE}(T_{\kappa_{min}}^{Poly}, f^{Poly}(\boldsymbol{\mu})) &= \Sigma_{00} + \varrho_0^2 + \kappa_{min}^2 \Sigma_{ii} - 2\kappa_{min}\Sigma_{i0} + \kappa_{min}^2 \varrho_i^2 - 2\kappa_{min}\varrho_0\varrho_1 \\ &= \Sigma_{00} + \varrho_0^2 + \kappa_{min}^2(\Sigma_{ii} + \varrho_i^2) - 2\kappa_{min}(\Sigma_{i0} + \varrho_0\varrho_1) \\ &= \Sigma_{00} + \varrho_0^2 - \tfrac{(\Sigma_{i0} + \varrho_0\varrho_i)^2}{(\Sigma_{ii} + \varrho_i^2)}. \end{aligned}$$

The latest result coincides with Troschke (2002, Equation 1.17).

Proof: The $\min_{\kappa} \mathrm{MSE}(T_\kappa^{Poly}, f^{Poly}(\boldsymbol{\mu}))$ was deduced by applying the following well-known three steps:

1. **Explicit calculation of the MSE$(T_\kappa^{Poly}, f^{Poly}(\boldsymbol{\mu}))$.**

$$\begin{aligned} \mathrm{MSE}(T_\kappa^{Poly}, f^{Poly}(\boldsymbol{\mu})) &= \kappa^2 \mathrm{var}(T_1^{Poly}) + (1-\kappa)^2 \mathrm{var}(T_2^{Poly}) \\ &+ 2\kappa(1-\kappa)\mathrm{cov}(T_1^{Poly}, T_2^{Poly}) + \kappa^2\iota_1^2 \\ &+ 2\kappa(1-\kappa)\iota_1\iota_2 + (1-\kappa)^2\iota_2. \end{aligned}$$

2. Differentiation. Common differential calculus

$$\begin{aligned}\frac{\partial \text{MSE}(T_\kappa^{Poly}, f^{Poly}(\boldsymbol{\mu}))}{\partial \kappa} &= 2\kappa \text{var}(T_1^{Poly}) - 2(1-\kappa)\text{var}(T_2^{Poly}) \\ &+ 2(1-2\kappa)\text{cov}(T_1^{Poly}, T_2^{Poly}) + 2\kappa\iota_1^2 \\ &+ 2(1-2\kappa)\iota_1\iota_2 - 2(1-\kappa)\iota_2,\end{aligned}$$

$$\frac{\partial^2 \text{MSE}(T_\kappa^{Poly}, f^{Poly}(\boldsymbol{\mu}))}{\partial^2 \kappa} = 2\text{var}(T_1^{Poly} - T_2^{Poly}) + 2(\iota_1 - \iota_2)^2 > 0.$$

3. Equating to zero. Setting $\frac{\partial \text{MSE}(T_\kappa^{Poly}, f^{Poly}(\boldsymbol{\mu}))}{\partial \kappa}$ to zero and solving the resulting linear equation for the unknown parameter κ, an optimal κ which minimises $\text{MSE}(T_{\kappa_{min}}^{Poly}, f^{Poly}(\boldsymbol{\mu}))$ can be obtained.

From $\frac{\partial \text{MSE}(T_\kappa^{Poly}, f^{Poly}(\boldsymbol{\mu}))}{\partial \kappa} = 0$ it follows:

$$\kappa_{min} = \frac{\text{var}(T_2^{Poly}) - \text{cov}(T_1^{Poly}, T_2^{Poly}) + \iota_2(\iota_2 - \iota_1)}{\text{var}(T_1^{Poly} - T_2^{Poly}) + (\iota_2 - \iota_1)^2},$$

which coincides with the Expression 5.1.4.

From the substitution of κ_{min} through Expression 5.1.4 in Equation 5.1.3 it follows:

$$\begin{aligned}\text{MSE}(T_{\kappa_{min}}^{Poly}, f^{Poly}(\boldsymbol{\mu})) &= \Sigma_{00} + \varrho_0^2 + \frac{(\Sigma_{i0}+\varrho_0\varrho_i)^2}{(\Sigma_{ii}+\varrho_i^2)^2}(\Sigma_{ii} + \varrho_i^2) - 2\frac{\Sigma_{i0}+\varrho_0\varrho_i}{\Sigma_{ii}+\varrho_i^2}(\Sigma_{i0} + \varrho_0\varrho_i) \\ &= \Sigma_{00} + \varrho_0^2 + \frac{(\Sigma_{i0}+\varrho_0\varrho_i)^2}{(\Sigma_{ii}+\varrho_i^2)} - 2\frac{(\Sigma_{i0}+\varrho_0\varrho_i)^2}{(\Sigma_{ii}+\varrho_i^2)} \\ &= \Sigma_{00} + \varrho_0^2 - \frac{(\Sigma_{i0}+\varrho_0\varrho_i)^2}{(\Sigma_{ii}+\varrho_i^2)}.\end{aligned}$$

■

Theorem 5.3 (Multinormal Distribution)

Assume that $\boldsymbol{y} = Vec\boldsymbol{X}'$ *is multinormally distributed with* $\text{E}[\boldsymbol{y}] = \boldsymbol{g}_* = \mathbf{1}_n \otimes \boldsymbol{\mu}$ *and* $\text{cov}(\boldsymbol{y}) = \boldsymbol{V}_* = \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}$, *then the resulting* κ_{min} *for the multinormal case is given by:*

$$\kappa_{min} = \frac{2\beta + n\alpha}{2\beta + (n-1)\alpha},$$

with $\gamma = \frac{1}{n}(2\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{a})'\boldsymbol{\Sigma}(2\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{a}),\quad \alpha = (tr\boldsymbol{A}\boldsymbol{\Sigma})^2$ *and* $\beta = tr(\boldsymbol{A}\boldsymbol{\Sigma})^2$.

Proof: See Appendix C.8.

5.2 Minimal MSE Estimation of NLSS Functions of Parameters of Univariate Distributions

Like in the multivariate case, T_3^{Poly} can be written as $T_\kappa^{Poly} = \kappa(T_1^{Poly} - T_2^{Poly}) + T_2^{Poly}$, with $\kappa = \frac{n}{n-1}$. In this case, the interest is also focused on finding an optimal κ or κ_{min}, so that $\text{MSE}(T_{\kappa_{min}}^{Poly}, f^{Poly}(\boldsymbol{\mu}))$ is minimised. $\text{MSE}(T_{\kappa_{min}}^{Poly}, f^{Poly}(\mu))$ is defined like in Theorem 5.2.

In Equation 5.1.4 the expression for κ_{min}, based on the approach introduced by Troschke (2002), is presented. Most of the expressions involved in this equation have already been deduced and presented in Theorem 3.4. In the following theorem the optimal expression for κ for the univariate case will be presented.

Theorem 5.4

Let $\boldsymbol{z} = \boldsymbol{y} - \text{E}[\boldsymbol{y}]$, with $\text{E}[\boldsymbol{z}] = \boldsymbol{0}$, $\boldsymbol{V}_* = \text{E}[\boldsymbol{z}\boldsymbol{z}'] = \text{cov}(\boldsymbol{y}) = \sigma^2 \boldsymbol{I}_n$ and $\boldsymbol{g}_* = E[\boldsymbol{y}] = \mu \mathbf{1}_n$. The resulting κ_{min} for the univariate case is given as follows:

$$\kappa_{min} = \frac{n[2A^2\mu\Phi + aA\Phi + A^2\Psi + (n-1)\alpha]}{(n-1)A^2\Psi + (2 + (n-1)^2)\alpha},$$

with $\gamma = \frac{\sigma^2}{n}(2A\mu + a)^2$.

Proof: See Appendix C.9.

Theorem 5.5 (Normal Distribution)

Assume that $\boldsymbol{x}$ is normally distributed with $\text{E}[\boldsymbol{x}] = \mu \mathbf{1}_n$ and $\text{cov}(\boldsymbol{x}) = \sigma^2 \boldsymbol{I}_n$. Furthermore, assume that $\boldsymbol{y} \sim N(\mu \mathbf{1}_n, \sigma^2 \boldsymbol{I}_n)$, with $\Phi = \Phi^N = 0$ and $\Psi = \Psi^N = 3\sigma^2$. The index N stands for representing the normal distribution.
The resulting κ_{min} for the normal case is given as follows:

$$\kappa_{min} = \frac{n+2}{n+1}.$$

Proof: See Appendix C.10.

Chapter 6

Application

Contents

In this section, asymptotical results for the estimation of NLSS from various references are compared with both the new estimator developed using the generalised Jackknife approach and the estimator obtained using the approach introduced by Troschke (2002).

The estimation quality or performance of the aforementioned estimators will be compared by means of their approximated and simulated MSEs for different distribution assumptions, parameter settings, correlation coefficients and sample sizes.

As first NLSS function, the ratio of means will be considered in Section 6.2.1. This function is a commonly used measure of comparison which can be used as an alternative to mean differences for analyzing continuous outcome variables. In the same section, estimation approaches for the ratio of means of two lognormal and gamma distributed variables are presented as special cases. In this specific context, Crow (1977) derived a Minimum

Variance Unbiased Estimator (MVUE) for the ratio of means of two independent lognormal distributions with equal and unequal shape parameters. In the same context, Shaban (1981) obtained estimators for the ratio of means of two independent lognormally distributed variables which are generally of smaller mean squared error than both the Maximum Likelihood (ML) and the MVUE as given in Crow (1977). For the gamma distribution, Crow (1977) also derived a MVUE for the ratio of means. In the same section the aforementioned estimators will be compared with those analysed and deduced in this work.

As second NLSS function the inverse of the mean will be considered in Section 6.3. For this function the ML estimator (one of the most frequently used estimators for this kind of functions) has no second and higher order moments. Srivastava and Bhatnagar (1981) presented a class of estimators free from the limitation of non-existence of moments and from the assumption of normality. Both estimators as well as their properties are presented in the same section. Further, both estimators will be compared with the estimators analysed and developed in this work.

Estimation of the odds in favour of an event will be considered in Section 6.4 as an example of a NLSS function for which the estimation approach deduced in this work can be applied. In the same section comparisons between estimators presented in literature and those analysed and developed in this work will also be carried out.

The following section introduces preliminaries, i.e. clarifying aspects and settings, needed for the simulation study.

6.1 Preliminaries

The aim of this preliminary section is to introduce the main concepts necessary for the simulation study to be carried out in this work. It consists basically of a general method for the generation of two correlated random variables is also presented as well as their higher order moments.

6.1.1 Generation of two Correlated Random Variables

After an intensive literature review in this topic, it could be concluded that with existing methods for generating correlated random variables it is quite simple, at least theoretically, to generate correlated random variables under the assumption of a particular distribution. In this respect, Fitzgerald et al. (2006) presented a technique for generating random variables that match moments and autocorrelations from a particular empirical distribution. They also showed that their approach is more accurate than traditional techniques.

Additional to the problem of assumption of a particular distribution, the setting of the correlation level between the variables to be generated is not straightforward. In this respect, Kleijnen (1974) derived different procedures for sampling two variables which have a predetermined correlation coefficient and possibly have prespecified marginal distributions. In his approach, instead of specifying a particular value for the correlation coefficient, the linear correlation may be maximized.

Another work related to this topic was written by Förster (1997), who describes a method for the Monte Carlo simulation of two correlated random variables. The author analysed linear combinations of stochastically independent random variables that are uniformly distributed over the interval $[0, 1]$. If a suitable matrix of coefficients is chosen, the subsequent transformation results in random variables with the desired distribution properties and covariance between the involved variables, with covariance between the involved variables X_1 and X_2 in the interval $[-0.3, 0.4]$.

In this work, the intention is to compare the performance of the estimators developed in the preceding sections. For this comparison, variables from a desired distribution function with different prespecified correlation levels, $\rho \in [-1, 1]$ will be generated.

The aforementioned approaches are not sufficient for the intended task. Therefore, it is necessary to develop a new approach for the generation of two correlated random variables with an arbitrary user-defined correlation level.

This approach was developed through personal communications with D. Trenkler and is presented as follows:

1. A bivariate i.i.d random sample $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ with

   $\mathrm{E}[\boldsymbol{x}_i] = \boldsymbol{0}$ and $\mathrm{cov}(\boldsymbol{x}_i)$=$\mathrm{E}[\boldsymbol{x}_i \boldsymbol{x}_i'] = \boldsymbol{I}_2$, for $i = 1, \cdots, n$ is generated.

   The random vector $\boldsymbol{x}_i$ is given by: $(X_{i1}, X_{i2})'$.

   Furthermore, $\boldsymbol{\mathcal{X}} = (X_1, X_2)'$ is a vector containing the two uncorrelated random variables.

   It holds that for suitable settings of parameters the following transformation

   $$\boldsymbol{\mathcal{Y}} = (Y_1, Y_2)' := \boldsymbol{C}\boldsymbol{B}\boldsymbol{\mathcal{X}} + \boldsymbol{b} \tag{6.1.1}$$

   yields a vector of correlated random variables, $\boldsymbol{\mathcal{Y}}$, with $\mathrm{E}[Y_j] = \mu_j$, $\mathrm{var}(Y_j) = \sigma_j^2$, $j = 1, 2$ and $\mathrm{corr}[Y_1, Y_2] = \rho$.

   Legal settings of these parameters are: $\mu_j \in \mathbb{R}$, $\sigma_j \in \mathbb{R}^+$ and $\rho \in [-1, 1]$.

   Furthermore, it is stated that:

   $$\boldsymbol{b} = \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} 1 & q \\ q & 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{C} = \begin{pmatrix} \tau_1 & 0 \\ 0 & \tau_2 \end{pmatrix},$$

   where q, τ_1 and τ_2 are given below.

2. Similarly, the following transformation is applied:

   $\boldsymbol{u} = (U_1, U_2)' := \boldsymbol{B}\boldsymbol{\mathcal{X}} = (X_1 + qX_2, qX_1 + X_2)'$, with

   $$\mathrm{var}(\boldsymbol{u}) = \boldsymbol{B}\boldsymbol{B}' = \begin{pmatrix} 1 + q^2 & 2q \\ 2q & 1 + q^2 \end{pmatrix},$$

and

$$\text{corr}[U_1, U_2] = \rho = \tfrac{2q}{(1+q^2)}.$$

From these equations it follows:

$$q = \begin{cases} \frac{1-\sqrt{1-\rho^2}}{\rho} & \text{for} \quad \rho \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

3. Now the transformation $\boldsymbol{v} = (V_1, V_2)' := \boldsymbol{Cu}$ is considered, with:

$$\text{var}(V_1) = \sigma_1^2 = \text{var}(\tau_1 U_1) = \tau_1^2 \text{var}(qX_1 + X_2) = \tau_1^2(1+q^2), \text{ and}$$

$$\text{var}(V_2) = \sigma_2^2 = \text{var}(\tau_2 U_1)\tau_2^2 \text{var}(X_1 + qX_2) = \tau_2^2(1+q^2).$$

These expressions yield:

$$\tau_1 = \sqrt{\frac{\sigma_1^2}{1+q^2}} \quad \text{and} \quad \tau_2 = \sqrt{\frac{\sigma_2^2}{1+q^2}}.$$

4. Finally, from transformation (6.1.1) the following vector of correlated random variables $\boldsymbol{\mathcal{Y}}$ can be obtained:

$$\boldsymbol{\mathcal{Y}} = (Y_1, Y_2)' = \boldsymbol{CB\mathcal{X}} + \boldsymbol{\mu} = \begin{pmatrix} \tau_1 X_1 + \tau_1 q X_2 + \mu_1 \\ \tau_2 q X_1 + \tau_2 X_2 + \mu_2 \end{pmatrix}. \tag{6.1.2}$$

Moments of two Correlated Variables

In this section, the third and fourth central moments, necessary for the computation of the covariance between T_1^{Poly} and T_2^{Poly} as well as their separated variances when the involved random variables are correlated, are presented.

Suppose $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ represents a realisation of an i.i.d. bivariate random sample with mean $\mathrm{E}[\boldsymbol{x}_i] = \mathbf{0}$ and covariance matrix $\mathrm{E}[\boldsymbol{x}_i\boldsymbol{x}_i'] = \boldsymbol{I}_2$. From properties of the variance it follows that $\mathrm{E}[X_j^2] = (\mathrm{E}[X_j])^2 + \mathrm{var}[X_j] = 1$, see e.g. Mood, Graybill and Boes (1974).

Let $\boldsymbol{Y} \in \mathbb{R}^{n\times 2}$ represent the matrix resulting from the transformation (6.1.2), with n the sample size and 2 the number of correlated variables. It is to allude that $\boldsymbol{Y}$ has a similar structure as $\boldsymbol{X}$ presented in Remark 6.1, so that a correlated random sample is given as follows:

$$\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n, \text{ with } \boldsymbol{y}_i = (Y_{i1}, Y_{i2})' \text{ and } \mathrm{E}[\boldsymbol{y}_i] = \boldsymbol{\mu} \text{ and } \mathrm{E}[(\boldsymbol{y}_i - \boldsymbol{\mu})(\boldsymbol{y}_i - \boldsymbol{\mu})'] = \boldsymbol{\Sigma}.$$

Similarly to Section 3.1 the random vectors $\boldsymbol{z}_i = \boldsymbol{y}_i - \boldsymbol{\mu}$, $i = 1, \ldots, n$, with existing moments $\mathrm{E}[\boldsymbol{z}_i] = \mathbf{0}$, $\mathrm{E}[\boldsymbol{z}_i\boldsymbol{z}_i'] = \boldsymbol{\Sigma}$, $\mathrm{E}[\boldsymbol{z}_i \otimes \boldsymbol{z}_i\boldsymbol{z}_i'] = \boldsymbol{\Phi}$ and $\mathrm{E}[\boldsymbol{z}_i\boldsymbol{z}_i' \otimes \boldsymbol{z}_i\boldsymbol{z}_i'] = \boldsymbol{\Psi}$ are considered.

Now, let $\boldsymbol{y} = \mathrm{Vec}(\boldsymbol{Y}')$ and $\boldsymbol{z} = \boldsymbol{y} - \mathrm{E}[\boldsymbol{y}]$, with $\mathrm{E}[\boldsymbol{y}] = \mathbf{1}_n \otimes \boldsymbol{\mu}$ and $\mathrm{cov}(\boldsymbol{y}) = \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}$, from this it follows that $\mathrm{E}[\boldsymbol{z}] = \mathbf{0}$ and $\mathrm{E}[\boldsymbol{z}\boldsymbol{z}'] = \mathrm{cov}(\boldsymbol{y}) = \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}$.

Notice that $\boldsymbol{z}$ decomposes into independent subvectors $\boldsymbol{z}_i$ with dimension $p = 2$, and that the elements of $\boldsymbol{z}_i$ (i.e. $\boldsymbol{z}_i = (z_{i1}, z_{i2})'$) are **correlated**.

To facilitate notation for the particular bivariate case the following vectors containing random variables are considered, i.e. $\boldsymbol{\mathcal{X}} = (X_1, X_2)'$, $\boldsymbol{\mathcal{Y}} = (Y_1, Y_2)'$ and $\boldsymbol{\mathcal{Z}} = (Z_1, Z_2)'$.

In Appendix A.5.2 the non-central as well as the central moments of two uncorrelated random variables, i.e. $\boldsymbol{\mathcal{X}} = (X_1, X_2)'$ have been presented.

In the statistical literature the third and fourth non-central moments of the most commonly used probability distributions have widely been reported, see for example Mood, Graybill and Boes (1974). The first four central moments of the correlated variables $\boldsymbol{\mathcal{Y}} = (Y_1, Y_2)'$ are presented in Remark A.6.

6.2 Estimation of the Ratio of Means (RM)

When a group of population means is given, one of the questions that normally arises is how sample information can be used to distinguish between these means. Many times the investigator starts analysing all the differences between each pair of means or expressing these means as linear combinations. Nevertheless, in some situations the knowledge about the group of means is obtained in a more natural way through the ratios between each pair of means. This concept is presented more formally in the following definition.

Definition 9 (Ratio of Means)

Let $\boldsymbol{x}_1 = (X_{11}, \cdots, X_{n1})'$ *and* $\boldsymbol{x}_2 = (X_{12}, \cdots, X_{n2})'$ *be two random samples drawn from some distribution with means* μ_1 *and* μ_2*, respectively. The* Ratio of Means, *read as the ratio of the mean* μ_2 *to the mean* μ_1*, is defined as:*

$$f(\boldsymbol{\mu}) = \mu_2/\mu_1,$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2)'$ *and* $\mu_1 \neq 0$.

In most applications the case $\mu_1 > 0$ and $\mu_2 > 0$ is of particular interest.

The ratio of means can be used as a measure of comparison, e.g. in a test of equivalence where the principal interest is to demonstrate whether two treatments differ more than a certain quantity (specified clinically). Hence, the equivalence is defined in terms of the ratio of means (Hauschke et al., 1999a,b).

Another example is the Consumer Price Index (CPI), which is a ratio of costs of purchasing a fixed set of items for two points in time. Sociologists are interested in measures like the ratio of the total monthly food budget to the total monthly income per family or the ratio of the total number of children to the total number of people residing in the household.

However, it has to be pointed out that statistical inference is more complicated for a ratio of parameters than for linear combinations of them. One of the difficulties in dealing with ratios arises in computing the variance of their estimators. In epidemiological studies, there are different factors which can be quantified using absolute measures, such as the risk difference, or by applying relative measures such as the relative risk or odds ratio. Both the relative risk and the odds ratio require more caution from an inferential point of view than the simple risk difference.

Beyene and Moineddin (2005) termed the ratio of population means as "a measure designed to quantify and benchmark the degree of relative concentration of an activity in the analysis of area localisation, which has received considerable attention in the geographic and economics literature as well as in the context of population health to quantify and compare health outcomes across spatial domains". They also pointed out that one commonly observed limitation of the ratio of population means is its widespread use as only a point estimation without an accompanying confidence interval.

A new approach for the point estimation of the ratio of means, free from the limitation of non-existence of moments and from the assumption of normality, is presented as follows.

6.2.1 Estimation of the Ratio of Means of Arbitrary Distributions

Suppose a bivariate i.i.d random sample $\boldsymbol{x}_1 = (X_{11}, X_{12})', \boldsymbol{x}_2 = (X_{21}, X_{22})', \ldots, \boldsymbol{x}_n = (X_{n1}, X_{n2})'$ from a 2–dimensional probability distribution, with means $\mathrm{E}[X_1] = \mu_1$ and $\mathrm{E}[X_2] = \mu_2$ and variances $\mathrm{var}(X_1) = \sigma_1^2$ and $\mathrm{var}(X_2) = \sigma_2^2$ is given. The multivariate mean and covariance matrix can be represented as follows:

$$E[\boldsymbol{x}_i] = \boldsymbol{\mu} = \begin{pmatrix} \mathrm{E}[X_1] = \mu_1 \\ \mathrm{E}[X_2] = \mu_2 \end{pmatrix}, \text{ and}$$

$$E[(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})'] = \boldsymbol{\Sigma} = \begin{pmatrix} \mathrm{var}(X_1) = \sigma_1^2 & \mathrm{cov}[X_1, X_2] = 0 \\ \mathrm{cov}[X_2, X_1] = 0 & \mathrm{var}(X_2) = \sigma_2^2 \end{pmatrix}.$$

The structure of the corresponding *bivariate sample matrix*, based on the sample matrix introduced in Definition 6, is presented in the following remark.

Remark 6.1 (Bivariate Sample Matrix)

Suppose $p = 2$ random variables X_1 and X_2 with n observations taken on each of the variables are given. The corresponding $n \times 2$ sample matrix, is given by:

$$\boldsymbol{X} = \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \\ \vdots & \vdots \\ X_{n1} & X_{n2} \end{pmatrix},$$

$\boldsymbol{X}$ can also be represented as:

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_n \end{pmatrix}, \text{ with } \boldsymbol{x}_i = \begin{pmatrix} X_{i1} \\ X_{i2} \end{pmatrix}', \text{ or as: } \boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix}', \text{ with } \boldsymbol{x}_j = \begin{pmatrix} X_{1j} \\ X_{2j} \\ \vdots \\ X_{nj} \end{pmatrix},$$

for $i = 1, \cdots, n$ and for $j = 1, 2$.

$\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ represents a realisation of an i.i.d. random sample from a bivariate (2–dimensional) distribution, with mean $\boldsymbol{\mu} = E[\boldsymbol{x}_i]$ and covariance matrix $\boldsymbol{\Sigma} = E[(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})']$.

Now, suppose that the ratio of the mean μ_2 to the mean μ_1, see Definition 9, has to be estimated.

In Section 2.2.3 a general estimation approach that can be applied for the statistical inference of functions of parameter(s) of a particular distribution has been presented. The general procedure has been summarised in Section 2.3. There it has been stated that the only condition for the application of the general estimation approach is that of smoothness of the function to be estimated, i.e. the function has to be of at least class C^2, see Definition 1.

In what follows it will be investigated, whether the function $f(\boldsymbol{\mu}) = \mu_2/\mu_1$ fulfills the smoothness condition, i.e. whether it is of at least class C^2.

Remark 6.2 (Smoothness Condition)

Let $f(\boldsymbol{\mu}) = \mu_2/\mu_1$. The first and second order derivatives of $f(\boldsymbol{\mu})$ are given by:

$$f^{(1)}(\boldsymbol{\mu}) = \begin{pmatrix} \frac{\partial f(\boldsymbol{\mu})}{\partial \mu_1} \\ \\ \frac{\partial f(\boldsymbol{\mu})}{\partial \mu_2} \end{pmatrix} = \begin{pmatrix} -\frac{\mu_2}{\mu_1^2} \\ \\ \frac{1}{\mu_1} \end{pmatrix}, \textit{ the gradient of the function } f(\boldsymbol{\mu}), \textit{ and}$$

$$f^{(2)}(\boldsymbol{\mu}) = \begin{pmatrix} \frac{\partial}{\partial \mu_1}(\frac{\partial f(\boldsymbol{\mu})}{\partial \mu_1}) & \frac{\partial}{\partial \mu_1}(\frac{\partial f(\boldsymbol{\mu})}{\partial \mu_2}) \\ \\ \frac{\partial}{\partial \mu_2}(\frac{\partial f(\boldsymbol{\mu})}{\partial \mu_1}) & \frac{\partial}{\partial \mu_2}(\frac{\partial f(\boldsymbol{\mu})}{\partial \mu_2}) \end{pmatrix} = \begin{pmatrix} \frac{2\mu_2}{\mu_1^3} & -\frac{1}{\mu_1^2} \\ \\ -\frac{1}{\mu_1^2} & 0 \end{pmatrix}, \textit{ the Hessian matrix.}$$

For further details see Definition 4.

It can be seen that the smoothness condition of function $f(\boldsymbol{\mu})$ is fulfilled.

For the estimation of NLSS functions, in this work, two estimation approaches have been investigated, i.e. $f(\hat{G}(\boldsymbol{X}))$ and $\hat{g}(f(\boldsymbol{x}_i), i = 1, \cdots, n)$, see Definition 8.

For the estimation of $f(\boldsymbol{\mu}) = \mu_2/\mu_1$ these estimation approaches investigated in this work are given by:

$$T_1 = f(\hat{G}(\boldsymbol{X})) = f(\bar{\boldsymbol{x}}) = \bar{x}_2/\bar{x}_1, \text{ and}$$

$$T_2 = \hat{g}(f(\boldsymbol{x}_i), i = 1, \cdots, n) = \tfrac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{x}_i) = \tfrac{1}{n}\sum_{i=1}^{n}(X_{i2}/X_{i1}).$$

As second step of the estimation approach developed in this work, the generalised Jackknife approach, see Section A.1.3, is used in order to generate an unbiased estimator for $f(\boldsymbol{\mu})$. This estimator is given by:

$$T_3 = \frac{T_1 - RT_2}{1 - R} = \frac{\left(\bar{x}_2/\bar{x}_1 - R\frac{1}{n}\sum_{i=1}^{n}(X_{i2}/X_{i1})\right)}{1 - R}, \text{ with } R = \frac{\mathrm{b}(T_1, f(\boldsymbol{\mu}))}{\mathrm{b}(T_2, f(\boldsymbol{\mu}))}.$$

In order to obtain an optimal estimator a linear adjustment as introduced by Troschke (2002) is applied. This estimator may be biased, but has smaller MSE than its unbiased counterpart generated by the generalised Jackknife approach.

For the calculation of the estimators T_3 and $T_{\kappa_{min}}$ the first and higher order moments of the estimators T_1 and T_2 are needed. They will be approximated by using the approach presented in Neudecker and Trenkler (2005a) and improved in the previous sections of this work.

For the particular case of the estimation of $f(\boldsymbol{\mu}) = \mu_2/\mu_1$, an approximation on the basis of a linear plus quadratic function, of the function itself as well as of the estimators $T_1 = f(\bar{\boldsymbol{x}})$ and $T_2 = \frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{x}_i)$, as made in Remark 3.1, is presented as follows:

$$f^{Poly}(\boldsymbol{\mu}) = a_0 + \boldsymbol{a}'\boldsymbol{\mu} + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu},$$

$$\begin{aligned} T_1^{Poly} = f^{Poly}(\bar{\boldsymbol{x}}) &= a_0 + \tfrac{1}{n}\boldsymbol{a}'\boldsymbol{X}'\mathbf{1}_n + \tfrac{1}{n^2}\mathbf{1}_n'\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}'\mathbf{1}_n \\ &= a_0 + \tfrac{1}{n}\boldsymbol{a}'\boldsymbol{X}'\mathbf{1}_n + \tfrac{1}{n^2}tr\mathbf{1}_n\mathbf{1}_n'\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}' \\ &= a_0 + \boldsymbol{a}'\bar{\boldsymbol{x}} + \bar{\boldsymbol{x}}'\boldsymbol{A}\bar{\boldsymbol{x}}, \text{ and} \end{aligned}$$

$$T_2^{Poly} = \frac{1}{n}\sum_{i=1}^{n} f^{Poly}(\boldsymbol{x}_i) = a_0 + \boldsymbol{a}'\bar{\boldsymbol{x}} + \tfrac{1}{n}\Sigma_{i=1}^{n}\boldsymbol{x}_i'\boldsymbol{A}\boldsymbol{x}_i,$$

where a_0 is a constant, $\boldsymbol{a}' = [f^{(1)}(\boldsymbol{\mu})]'$ is the transpose of gradient vector and $\boldsymbol{A} = \frac{1}{2}f^{(2)}(\boldsymbol{\mu})$ the Hessian matrix of the function $f(\boldsymbol{\mu})$ divided by two, see Remark 6.2.

Remark 6.3

Notice that replacing $\boldsymbol{a} = f^{(1)}(\boldsymbol{\mu})$ and $\boldsymbol{A} = \frac{1}{2}f^{(2)}(\boldsymbol{\mu})$ by their respective expressions, given in Remark 6.2, the approximation $f^{Poly}(\boldsymbol{\mu})$ is equal to a_0. Suppose $a_0 = f(\boldsymbol{\mu})$ then $f^{Poly}(\boldsymbol{\mu}) = f(\boldsymbol{\mu})$.

In Section 3.1 it was pointed out that both approximated estimators T_1^{Poly} and T_2^{Poly} are biased for $f^{Poly}(\boldsymbol{\mu})$. The bias terms are presented in the following remark.

Remark 6.4 (Bias of Approximated Estimators T_1^{Poly} and T_2^{Poly})

Since $\boldsymbol{A\Sigma}$ *is given by:*

$$\boldsymbol{A\Sigma} = \frac{1}{2}\begin{pmatrix} \frac{2\mu_2}{\mu_1^3} & -\frac{1}{\mu_1^2} \\ \\ -\frac{1}{\mu_1^2} & 0 \end{pmatrix}\begin{pmatrix} \sigma_1^2 & 0 \\ \\ 0 & \sigma_2^2 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} \frac{2\mu_2\sigma_1^2}{\mu_1^3} & -\frac{\sigma_2^2}{\mu_1^2} \\ \\ -\frac{\sigma_1^2}{\mu_1^2} & 0 \end{pmatrix},$$

and the trace of $\boldsymbol{A\Sigma}$, *say* $tr\boldsymbol{A\Sigma}$, *is given by:* $\dfrac{\mu_2\sigma_1^2}{\mu_1^3}$, *consequently the bias terms of the approximated estimators* T_1^{Poly} *and* T_2^{Poly} *are given by:*

$$b(T_1^{Poly}, f^{Poly}(\boldsymbol{\mu})) = \tfrac{1}{n}tr\boldsymbol{A\Sigma} = \tfrac{\mu_2\sigma_1^2}{n\mu_1^3}, \text{ and}$$

$$b(T_2^{Poly}, f^{Poly}(\boldsymbol{\mu})) = tr\boldsymbol{A\Sigma} = \tfrac{\mu_2\sigma_1^2}{\mu_1^3}, \text{ see Definition 20.}$$

The expressions of the approximations, on the basis of a linear plus quadratic function, of the estimators to be investigated in this work corresponding to the whole sample $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ will be summarized in the following remark.

Remark 6.5

For the estimation of $f^{Poly}(\boldsymbol{\mu})$ *the following estimators are proposed:*

$$\begin{aligned} T_1^{Poly} &= f_0 + \boldsymbol{f}'\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{F}_1\boldsymbol{y} \\ T_2^{Poly} &= f_0 + \boldsymbol{f}'\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{F}_2\boldsymbol{y} \\ T_3^{Poly} &= f_0 + \boldsymbol{f}'\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{F}_3\boldsymbol{y} \\ T_{\kappa_{min}}^{Poly} &= \kappa_{min}(T_1^{Poly} - T_2^{Poly}) + T_2^{Poly}, \end{aligned}$$

where $\boldsymbol{y} = \mathrm{Vec}\boldsymbol{X}'$, f_0 *is a constant,* $\boldsymbol{f} = \frac{1}{n}\mathbf{1}_n \otimes \boldsymbol{a}$, $\boldsymbol{F}_1 = \frac{1}{n^2}(\mathbf{1}_n\mathbf{1}_n') \otimes \boldsymbol{A}$, $\boldsymbol{F}_2 = \frac{1}{n}\boldsymbol{I}_n \otimes \boldsymbol{A}$, $\boldsymbol{F}_3 = -\frac{1}{n(n-1)}\boldsymbol{L} \otimes \boldsymbol{A}$ *and* $\boldsymbol{L} = \boldsymbol{I}_n - \mathbf{1}_n\mathbf{1}_n'$. *Furthermore,* $\boldsymbol{a}$ *and* $\boldsymbol{A}$ *are given in Remark 6.2 and* κ_{min} *is given as in Theorem 5.1.*

The corresponding expressions for the covariance and variances of the aforementioned estimators have been presented in Section 3.1.1.

In this work, one of the principal objectives is to compare the MSE of the estimators T_1, T_2, T_3 and $T_{\kappa_{min}}$ of $f(\boldsymbol{\mu})$ under different distribution assumptions, parameter settings and for different sample sizes. However, it is also intended to compare the performance of the aforementioned estimators for correlated random variables, where the correlation level between those variables is assumed to be known. The general approach for generating two correlated random variables as well as their central moments is presented in Section 6.1.1.

The comparison of the MSEs will be done by means of their approximated expressions and by simulations in the following section.

Comparison of Estimators of the Ratio of Means

As stated in Section 1.2, most published methods concerning the estimation of functions of distribution parameters are asymptotic in nature and based on the assumption of uncorrelated standard normal random variables.

Since these assumptions are very restrictive, i.e. no correlation, standard normal distribution, asymptotic, in this work the interest is also concentrated on comparing the performance of the different estimators under different conditions, so that the following questions can be answered:

1. How does the sample size influence the performance of the estimators?
2. How do the estimators perform under the assumption of any arbitrary probability distribution with different parameter settings?
3. Does the parameter setting influence the performance of the estimators? In the same respect this raises the question: How do the estimators behave, when the mean of the variable in the denominator is close to zero?
4. Does the level of correlation between the variables involved influence the performance of the estimators?

It can be summarised as follows: The performance of the estimators T_1, T_2, T_3 and $T_{\kappa_{min}}$ will be compared by means of variables following different probability distribution assumptions, parameter settings, such as μ and σ^2, sample sizes n and correlation levels ρ (including $\rho = 0$).

Remark 6.6

In order to make a comparison of the performance of the estimators for two correlated random variables possible, a new approach for generating correlated random variables has been presented in Section 6.1.1.
For the application of this approach the generated random sample $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ has to fulfill the condition $\mathrm{E}[\boldsymbol{x}_i] = \mathbf{0}$ and $\mathrm{E}[(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})'] = \boldsymbol{I}_2$.
It is important to remember that: $q = \frac{1-\sqrt{1-\rho^2}}{\rho}$, for $\rho \neq 0$ and $q = 0$, for $\rho = 0$.

Assumed Distributions for the Estimation of the Ratio of Means

For concepts and definitions related to random variables and distribution functions refer to Mood, Graybill and Boes (1974).

Remark 6.7

The third and fourth non-central moments of the uncorrelated random variables $\mathrm{E}[X_j^3]$ and $\mathrm{E}[X_j^4]$, $j = 1, 2$ following a given distribution have widely been reported in the statistical literature, e.g. Mood, Graybill and Boes (1974). Those of this distributions assumed in the simulation study are given in Appendix A.1.4.

In order to make more generalisable comparisons between the estimators being analysed, two variables following different distributions will be generated. These distributions and their first four moments are presented as follows.

Distribution I: Standard Normal Distribution

Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ be a realisation of an i.i.d. bivariate random sample drawn from a standard normal distribution with parameters $\mathrm{E}[\boldsymbol{x}_i] = \mathbf{0}$ and $E[(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})'] = \boldsymbol{I}_2$. Since variables X_j, $j = 1, 2$ have a standard normal distribution, the assumption $\mathrm{E}[\boldsymbol{x}_i] = \mathbf{0}$ and $E[(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})'] = \boldsymbol{I}_2$ presented in Remark 6.6 is automatically fulfilled.

The first four central moments of the new variable generated using the approach presented in Section 6.1.1, i.e. Y_j, $j = 1, 2$, under standard normal distribution assumption of X_j, $j = 1, 2$, were obtained using the equations given in Remark A.6 and are presented as follows:

Central Moments of Y_j, $j = 1, 2$

$$\mathrm{E}[Y_j - \mu_j] = \mathrm{E}[Z_j] = 0$$

$$\mathrm{E}[(Y_j - \mu_j)^2] = \mathrm{E}[Z_j^2] = \tau_j^2(1 + q^2) = \sigma_j^2$$

$$\Phi_j = \mathrm{E}[(Y_j - \mu_j)^3] = \mathrm{E}[Z_j^3] = 0$$

$$\Psi_j = \mathrm{E}[(Y_j - \mu_j)^4] = \mathrm{E}[Z_j^4] = 3\tau_j^4[(1 + q^4) + 2q^2].$$

Distribution II: Exponential Distribution

Let X_j, $j = 1, 2$ be exponentially distributed with $\mathrm{E}[X_j] = \frac{1}{\lambda_j} = 1$ and $\mathrm{var}(X_j) = \frac{1}{\lambda_j^2} = 1$.

As following it will be examined whether the condition $\mathrm{E}[\boldsymbol{x}_i] = \mathbf{0}$ and $E[(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})'] = \boldsymbol{I}_2$ presented in Remark 6.6 is fulfilled for this distribution, i.e.:

$$E[\boldsymbol{x}_i] = \begin{pmatrix} \frac{1}{\lambda_1} \\ \frac{1}{\lambda_2} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \neq \mathbf{0}, \text{and } \mathrm{cov}(\boldsymbol{x}_i) = \begin{pmatrix} \mathrm{var}(X_1) & \mathrm{cov}(X_1 X_2) \\ \mathrm{cov}(X_1 X_2) & \mathrm{var}(X_2) \end{pmatrix} = \begin{pmatrix} \frac{1}{\lambda_1^2} & 0 \\ 0 & \frac{1}{\lambda_2^2} \end{pmatrix} = \boldsymbol{I}_2.$$

The condition is not fulfilled, therefore, it will be necessary to apply a linear transformation on the variables X_j, $j = 1, 2$; so that new variables, say $X_{j*} = X_j - a$ with $a \in \mathbb{R}$, have a shifted exponential distribution with probability density function given as presented in the following definition.

Let X be an exponential-distributed random variable with PDF given as above. Random variables with the form $X_* = X \pm a$ follow a particular distribution called *shifted exponential distribution* with *shift* parameter a.

The PDF of this particular distribution is presented in the following definition.

Definition 10 (Shifted Exponential Distribution)

Let X be an exponential-distributed random variable with parameter $\lambda > 0$.
A continuous random variable $X_* = X + a$ has a shifted exponential distribution, *with parameter $\lambda > 0$ and* shift *parameter $a \in \mathbb{R}$, if its probability density function (PDF) is given by:*

$$f_{X_*}(x_*) = \begin{cases} \lambda \exp[-\lambda(x_* - a)] & \text{if } x_* > a, \\ 0 & \text{otherwise} \end{cases}$$

The plot below shows the density function of an exponential random variable with parameter $\lambda = 1$ in black and of the shifted exponential random variable in red, with parameters $\lambda = 1$ and $a = 1$.

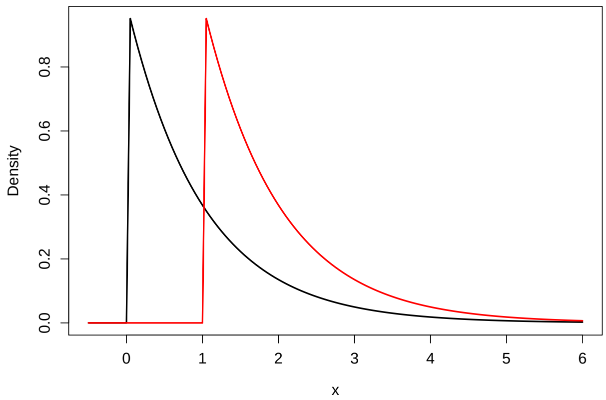

Figure 6.1: Exponential and shifted exponential distribution

The mean and variance of the variable X_* are given by:

$$\mathrm{E}[X_*] = \mathrm{E}[X - a] = \mathrm{E}[X] - a, \text{ and } \mathrm{var}(X_*) = \mathrm{var}(X - a) = \mathrm{var}(X).$$

In order to know whether the variables X_{j*}, $j = 1, 2$ with $\lambda = 1$ and shift parameter $a = 1$ fulfill the condition presented in Remark 6.6, the mean and variance of these variables will be presented as follows:

Mean and Variance of X_{j*}, $j = 1, 2$

$$\mathrm{E}[X_{j*}] = \mathrm{E}[X_j - 1] = \mathrm{E}[X_j] - 1 = 0$$

$$\mathrm{var}(X_{j*}) = \mathrm{var}(X_j - 1) = \mathrm{var}(X_j) = 1.$$

As can be seen the shifted exponential distributed variables X_{j*}, $j = 1, 2$ with parameter $\lambda = 1$ fulfill the condition presented in Remark 6.6.

The first four central moments of Y_j, $j = 1, 2$, under shifted exponential distribution assumptions of X_{j*}, $j = 1, 2$ X_j, $j = 1, 2$, with parameter $\lambda = 1$ are presented as follows:

Central Moments of Y_1 and Y_2

$$\mathrm{E}[Z_j] = 0$$

$$\mathrm{E}[Z_j^2] = \sigma_j^2$$

$$\Phi_j = \mathrm{E}[Z_j^3] = \tau_j^3(1 + q^3)$$

$$\Psi_j = \mathrm{E}[Z_j^4] = \tau_j^4[(1 + q^4) + 6q^2].$$

Distribution III: Uniform Distribution

Let X_j, $j = 1, 2$ be uniformly distributed in the interval $[a = -\sqrt{3}, b = \sqrt{3}]$ with mean and variance given as follows:

Mean and Variance of X_j, $j = 1, 2$

$$\mathrm{E}[X_j] = \tfrac{a_j + b_j}{2} = \tfrac{\sqrt{3} - \sqrt{3}}{2} = 0$$

$$\mathrm{var}(X_j) = \tfrac{(b_j - a_j)^2}{12} = \tfrac{(\sqrt{3} + \sqrt{3})^2}{12} = 1.$$

The variables X_j, $j = 1, 2$ also fulfill the condition presented in Remark 6.6.

The first four central moments of Y_j, $j = 1, 2$, under uniform distribution assumptions of X_j, $j = 1, 2$ X_j, $j = 1, 2$, with parameters $a = -\sqrt{3}$ and $b = \sqrt{3}$ are presented as follows:

Central Moments of Y_1 and Y_2

$$\mathrm{E}[Z_j] = 0$$

$$\mathrm{E}[Z_j^2] = \sigma_j^2$$

$$\Phi_j = \mathrm{E}[Z_j^3] = 0$$

$$\Psi_j = \mathrm{E}[Z_j^4] = \tau_j^4[1.8(1 + q^4) + 6q^2].$$

Assumed Parameter Settings (PS) for the Estimation of the Ratio of Means

In addition to the three distribution assumptions presented above, for the estimation of $f(\boldsymbol{\mu})$, the performance of the estimators T_1, T_2, T_3, and $T_{\kappa_{min}}$ will be compared for different parameter settings of the random variables Y_1 and Y_2.

These parameter settings are presented as follows:

- **PS 1: Equal Means and Variances** ($\mu_1 = \mu_2$ **and** $\sigma_1^2 = \sigma_2^2$)
  $\mu_1 = \mathrm{E}[Y_1] = 10$, $\mu_2 = \mathrm{E}[Y_2] = 10$, $\sigma_1^2 = \mathrm{var}(Y_1) = 1$, $\sigma_2^2 = \mathrm{var}(Y_2) = 1$
- **PS 2: Equal Means and Unequal Variances** ($\mu_1 = \mu_2$ **and** $\sigma_1^2 \neq \sigma_2^2$)
  $\mu_1 = \mathrm{E}[Y_1] = 10$, $\mu_2 = \mathrm{E}[Y_2] = 10$, $\sigma_1^2 = \mathrm{var}(Y_1) = 1$, $\sigma_2^2 = \mathrm{var}(Y_2) = 3$
- **PS 3: Unequal Means and Equal Variances** ($\mu_1 \neq \mu_2$ **and** $\sigma_1^2 = \sigma_2^2$)
  $\mu_1 = \mathrm{E}[Y_1] = 10$, $\mu_2 = \mathrm{E}[Y_2] = 20$, $\sigma_1^2 = \mathrm{var}(Y_1) = 1$, $\sigma_2^2 = \mathrm{var}(Y_2) = 1$
- **PS 4: Unequal Means and Variances** ($\mu_1 \neq \mu_2$ **and** $\sigma_1^2 \neq \sigma_2^2$)
  $\mu_1 = \mathrm{E}[Y_1] = 10$, $\mu_2 = \mathrm{E}[Y_2] = 20$, $\sigma_1^2 = \mathrm{var}(Y_1) = 1$, $\sigma_2^2 = \mathrm{var}(Y_2) = 3$
- **PS 5:** $\mu_1 \approx 0$ **and** $\sigma_1^2 = \sigma_2^2$
  $\mu_1 = \mathrm{E}[Y_1] = 1/2$, $\mu_2 = \mathrm{E}[Y_2] = 10$, $\sigma_1^2 = \mathrm{var}(Y_1) = 1$, $\sigma_2^2 = \mathrm{var}(Y_2) = 1$
- **PS 6:** $\mu_1 \approx 0$ **and** $\sigma_1^2 \neq \sigma_2^2$
  $\mu_1 = \mathrm{E}[Y_1] = 1/2$, $\mu_2 = \mathrm{E}[Y_2] = 10$, $\sigma_1^2 = \mathrm{var}(Y_1) = 1$, $\sigma_2^2 = \mathrm{var}(Y_2) = 3$.

Results from Approximated Expressions and Discussion

Numerical results of the approximated $\mathrm{MSE}(T_l, f(\boldsymbol{\mu}))$, i.e. $\mathrm{MSE}(T_l^{Poly}, f^{Poly}(\boldsymbol{\mu}))$, $l = 1, 2, 3, \kappa_{min}$ for different variable settings, i.e. standard normal, exponential, uniform distribution, $\rho = -0.9$, $\rho = -0.5$, $\rho = 0$, $\rho = 0.5$ and $\rho = 0.9$, $n = 10$, $n = 100$ and $n = 1200$ as well as the parameter settings presented above, are presented in Appendix D. Each table, i.e. Table D.1 - D.3, contains the numerical results for a different probability distribution. In the following discussion each table will be analysed separately. However, it is important to illustrate how the tables are to be read.

The tables contain the MSEs of four approximated estimators whose names are given across the top, i.e. T_1^{Poly}, T_2^{Poly}, T_3^{Poly} and $T_{\kappa_{min}}^{Poly}$. Each estimation is carried out with different sample sizes which are also given across the top of the table, i.e. $n = 10$, $n = 100$ and $n = 1200$. This accommodation on the table makes it possible to compare the approximated MSEs or performance of each estimator across the different sample sizes by moving 4 positions from the left to the right of the table and between estimators by moving 1 position from the left to the right of the table.

The 6 different parameter settings (PS), as given above, were accommodated as block rows, where each block row consists of 5 rows representing 5 different correlation levels ρ, including $\rho = 0$. In this way, the performance of each estimator for a given parameter setting and a given correlation level can be compared simultaneously. The top-down correlation levels in each PS block row represent the following correlations: high negative, medium negative, no correlation, medium positive and high positive.

An illustration by means of an arbitrary example is presented as follows:

Consider the value 1.22e-05 in the row number 6, i.e. first correlation row in the second block row, of Table D.1. This value shows the approximated MSE of estimator T_1, presented as T_1^{Poly}, calculated from a medium negative-correlated bivariate normal sample, with equal means and unequal variances, consisting of 10 elements.

Moving 4 positions to the right, it can be seen that the approximated MSE of the same estimator takes the value 1.9e-07, when the bivariate sample as described above consists of 100 elements instead of 10. Furthermore, moving another 4 positions to the right the approximated MSE of the same estimator takes the value $1.7\mathrm{e}{-09}$ when the bivariate sample consists of 1200 elements. From this comparison of the sample sizes it can be seen that the $\mathrm{MSE}(T_1^{Poly}, f^{Poly}(\boldsymbol{\mu}))$ decreases by at least two orders of magnitude as the sample size increases.

Now, keeping the same initial value 1.22e-05 and moving one row lower but staying in the same block row, i.e. from correlation level $\rho = -0.5$ to $\rho = 0$, it can be seen that the approximated MSE of the same estimator takes the value 1.12e-05. On the other hand, moving one block row lower and keeping the same correlation level, i.e. $\rho = -0.5$, the approximated MSE of the same estimator calculated from a medium negative-correlated bivariate normal sample, with unequal means and variances, consisting of 10 elements is shown.

The main findings from the results presented in the Tables D.1, D.2 and D.3 can be compared by keeping one or two variable settings fixed and varying the rest.

From Table D.1. Normal distribution:

- Keeping the PS and ρ fixed and varying n it can be seen that the approximated MSE of all estimators decreases by at least two orders of magnitude as the sample size increases. This happens for all PS and ρ values.
- Keeping n and PS fixed and varying ρ it can be seen that the approximated MSE of all estimators, except of T_2, decreases as the top-down correlation levels increase. This always happens for large sample sizes. For small sample sizes it only happens for PS 1,4 and 6, while for PS 2,3 and 5 it seems that the correlation does not play a big role.
- Keeping n and ρ fixed and varying PS it can be seen that the approximated MSE of all estimators for PS 5 and 6 are bigger, while no significant differences for the first four PS are observed.
- Keeping n, ρ and PS fixed and varying T_l^{Poly}, for $l = 1, 2, 3, \kappa_{min}$ it can be seen that the smallest approximated MSEs are obtained with the estimator $T_{\kappa_{min}}$, the second smallest with T_3, the third with T_1 and the highest with T_2. For higher sample sizes, the approximated MSE of estimators $T_{\kappa_{min}}$ and T_3 are almost identical.

From Table D.2. Exponential distribution:

- Keeping the PS and ρ fixed and varying n it can be seen that the approximated MSE of all estimators decreases by two orders of magnitude as the sample size increases. For PS 5 and 6 the decrease is by five orders of magnitude when the sample size increases from 10 to 100.

- Keeping n and PS fixed and varying ρ it can be seen that the approximated MSEs of all estimators decreases as the top-down correlation levels increase. This always happens for large sample sizes. For small sample sizes it only happens for PS 1-4.
- Keeping n and ρ fixed and varying PS it can be seen that as exposed for the normal distribution, see Table D.1, the approximated MSE of all estimators for PS 5 and 6 are much bigger, while no difference is observed among the first four PS.
- Keeping n, ρ and PS fixed and varying T_l^{Poly}, for $l = 1, 2, 3, \kappa_{min}$ it can be seen that, as for the normal distribution, the smallest approximated MSEs are obtained with the estimator $T_{\kappa_{min}}$, the second smallest with T_3, the third with T_1 and the highest with T_2. For higher sample sizes, the approximated MSE of estimators $T_{\kappa_{min}}$ and T_3 are almost identical.

From Table D.3. Uniform distribution:

- Keeping the PS and ρ fixed and varying n it can be seen that the approximated MSE of all estimators decreases by two orders of magnitude as the sample size increases. As obtained for the normal distribution, this happens for all PS and ρ values.
- Keeping n and PS fixed and varying ρ it can be seen that the approximated MSE of all estimators decreases as the top-down correlation levels increase. Larger positive correlation level ρ leads to a smaller approximated MSEs in many parameter settings. This happens for all sample sizes.
- Keeping n and ρ fixed and varying PS it can be seen that as exposed for the two preceding distributions, see Table D.1 and D.2, the approximated MSE of all estimators for PS 5 and 6 are bigger, while no difference is observed among the first four PS.
- Keeping n, ρ and PS fixed and varying T_l^{Poly}, for $l = 1, 2, 3, \kappa_{min}$ it can be seen that, as for the normal and exponential distribution, the smallest approximated MSEs are obtained with the estimator $T_{\kappa_{min}}$, the second smallest with T_3, the third with T_1 and the highest with T_2. For higher sample sizes, the performances of estimators $T_{\kappa_{min}}$ and T_3 are almost identical.

Now, fixing all parameters and varying the probability distribution and T_l^{Poly}, for $l = 1, 2, 3, \kappa_{min}$ the following conclusions can be drawn:

- For all distributions where the means and the variances are equal, i.e. PS 1, the approximated MSEs are smaller than those for the remaining cases. The largest approximated MSEs are obtained when $\mathrm{E}[Y_1]$ is near 0, i.e. PS 5 and 6. This can be explained from the fact that small changes on the $\mathrm{E}[Y_1]$, when it is close to zero, represent big changes on the MSE of all estimators.
- For all distributions, T_2 has higher approximated MSEs than the rest of the estimators, especially when $\mathrm{E}[Y_1]$ is near 0. Conversely, it can be seen that the approximated MSE of T_2 does not improve meaningfully as the sample size increases. For more details see Rao (1952) where the estimator T_2 was considered as inconsistent.
- For all distributions, $\mathrm{MSE}(T_3^{Poly}, f^{Poly}(\boldsymbol{\mu}))$ is slightly smaller than $\mathrm{MSE}(T_1^{Poly}, f^{Poly}(\boldsymbol{\mu}))$. This speaks in favour of the estimator T_3, which is a consistent estimator with small MSE, despite being a linear combination of a consistent and an inconsistent estimator, i.e. a linear combination of T_1 and T_2. However, the inconsistent part decreases with increasing sample size.
- For all distributions, the smallest approximated MSEs are obtained with the estimator $T_{\kappa_{min}}$. It demonstrates that the application of the *linear adjustment* presented in Lemma 5.1 has made a notable improvement in $\mathrm{MSE}(T_3^{Poly}, f^{Poly}(\boldsymbol{\mu}))$.
- The underlying distribution does not seem to play a big role in the performance of the different estimators. The estimators are highly accurate irrespective of the underlying distribution. A difference between the performance of the estimators could only be observed for parameter settings 4 and 5, where the approximated MSEs of the estimators under assumption of exponential distribution were higher.

From the above findings, $T_{\kappa_{min}}$ can be seen to be the best among the presented estimators, for all distributions, despite being slightly biased. T_3 is also highly accurate for the estimation of the ratio of means. T_2 cannot be recommended, whereas T_1 for larger sample size can be recommended in most cases.

Results from Simulations and Discussion

For the estimated or simulated MSEs of the estimators being compared in this section, 5000 repetitions, say *Rep*, will be carried out. For each repetition the squared error, i.e. the squared difference between each estimator and the function $f(\boldsymbol{\mu})$, is calculated and then the average of those repetitions is computed, see Definition 22, i.e. $\widehat{\text{MSE}}(T_l, f(\boldsymbol{\mu})) = \sum_{i=1}^{Rep} \frac{(T_{li}-f(\boldsymbol{\mu}))^2}{Rep}$, for $l = 1, 2, 3, \kappa_{min}$.

Numerical results of $\widehat{\text{MSE}}(T_l, f(\boldsymbol{\mu}))$, $l = 1, 2, 3, \kappa_{min}$ for the same variable settings as presented above are presented in tables in Appendix E. Each table, i.e. Table E.1 - E.3, contains the numerical results for a different probability distribution. The tables can be read in the same way as those presented above.

Analysing each table separately similar results are obtained as above. However, some differences are outlined as follows:

From Table E.1 - E.3 (Normal, Exponential and Uniform distributions) it can be seen that:

- The simulated or estimated MSEs are much higher than the approximated MSEs for all different variable settings. Those differences are by at least three orders of magnitude.
- The estimated variance of the estimator T_2 tends to infinity, when μ_1 approaches zero.
- Keeping n and PS fixed and varying ρ the estimated MSE of all estimators decreases as the top-down correlation levels increase. For small sample sizes this only happens for PS 1 - 4.

From the simulated results, $T_{\kappa_{min}}$ has also the smallest MSE among the presented estimators, for all distributions, despite being slightly biased. This represents a performance improvement on the estimation of NLSS functions, such as the ratio of means.

In the next section, the performance of the estimators analysed in this work will be compared, as made in this section by means of simulations, with existing solutions from literature where an specific distribution is assumed.

6.2.2 Comparisons with Existing Solutions

As mentioned in Section 2.1, Crow (1977) derived a Minimum Variance Unbiased Estimator (MVUE) for the ratio of means of two independent lognormal distributions with equal and unequal shape parameters. His unbiased estimator was used to evaluate the effect of seeding. He compared the ratio of means of seeded precipitation to the mean of natural precipitation. Initially the shape parameter σ was assumed to be unchanged, but a MVUE for the more general situation in which the shapes differ is also given. In this respect, Shaban (1981) obtained estimators for the ratio of means of two independent lognormal distributed variables which are generally of smaller mean squared error than both the Maximum Likelihood (ML) and the MVUE as given in Crow (1977). For the gamma distribution, Crow (1977) also derived a MVUE for the ratio of means. These approaches are presented in Appendix B.

In this section, the estimation approach developed in this work, applicable for any arbitrary probability distribution, will be compared with specific estimators for lognormal and gamma probability distributions, by means of simulations.

Estimation of the RM of two Uncorrelated Lognormal Variables

Suppose two independent random variables X_1 and X_2 distributed according to a lognormal distribution, with means $\mathrm{E}[X_1] = \exp[\mu_1 + \sigma_1^2/2]$ and $\mathrm{E}[X_2] = \exp[\mu_2 + \sigma_2^2/2]$ and variances $\mathrm{var}(X_1) = [\exp[\sigma_1^2] - 1]\exp[2\mu_1 + \sigma_1^2]$ and $\mathrm{var}(X_2) = [\exp[\sigma_2^2] - 1]\exp[2\mu_2 + \sigma_2^2]$, are given, so that $U_1 = \log(X_1)$ and $U_2 = \log(X_2)$ are normally distributed with parameters (μ_1, σ_1^2) and (μ_2, σ_2^2).

Estimators of the expected value and variance of the variables U_1 and U_2 are given by:

$$\bar{U}_1 = \frac{1}{n}\sum_{i=1}^{n} U_{1i}, \bar{U}_2 = \frac{1}{n}\sum_{j=1}^{n} U_{2j}, S_{U_1}^2 = \frac{1}{n}\sum_{i=1}^{n}(U_{1i} - \bar{U}_1)^2 \text{ and } S_{U_2}^2 = \frac{1}{n}\sum_{i=1}^{n}(U_{1i} - \bar{U}_2)^2,$$

where U_{1i} and U_{2j}, for $i = 1, \cdots, n$ and $j = 1, \cdots, n$ are the elements of the normal distributed samples $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$, respectively, i.e. $\boldsymbol{u}_1 = (U_{11}, U_{12}, \cdots, U_{1n})'$ and $\boldsymbol{u}_2 = (U_{21}, U_{22}, \cdots, U_{2n})'$.

Different properties of the lognormal distribution are presented in Remark B.1.

Shaban (1981) derived an estimator of $\varrho = \exp[a(\mu_2 - \mu_1) + b(\sigma_2^2 - \sigma_1^2)]$, see Equation B.1.4, with smaller MSE than the ML and the MVUE deduced by Crow (1977), for the following three situations:

- **1st Situation:** The shape parameters are equal $\sigma_1^2 = \sigma_2^2 = \sigma^2$ and σ^2 known.
- **2nd Situation:** The shape parameters are equal $\sigma_1^2 = \sigma_2^2 = \sigma^2$ and σ^2 unknown.
- **3rd Situation:** $\sigma_1^2 \neq \sigma_2^2$ and both are unknown.

The performance of the estimators presented in Remark 6.5 will be compared with those estimators presented in Shaban (1981), the ML estimator and the MVUE deduced by Crow (1977) for the three aforementioned situations.

These estimators are presented for the three situations listed above in Appendix B.1.

In the following remark, it is shown how the function to be estimated looks like for fixed parameter values a and b and for the aforementioned situations.

Remark 6.8

Assume that in $\varrho = \exp[a(\mu_2 - \mu_1) + b(\sigma_2^2 - \sigma_1^2)]$, $a = 1$ and $b = 1/2$, then it follows:

(i) *if $\sigma_1^2 = \sigma_2^2$, 1st and 2nd Situation from above, then the function* $\varrho = \exp[(\mathrm{E}[U_2]]/\exp[(\mathrm{E}[U_1])] = \exp[\mu_2 - \mu_1]$,

(ii) *if $\sigma_1^2 \neq \sigma_2^2$, 3rd Situation from above, then the function $\varrho = \mathrm{E}[X_2]/\mathrm{E}[X_1]$.*

Since the function $\varrho = \exp[\mu_2 - \mu_1]$ is the same for the first and second situation from above, the comparisons by means of simulations with different parameter settings can be made for these two situations simultaneously.

The parameter settings to be used in the simulation study and the results from the comparisons are presented as follows.

Assumed Parameter Settings (PS) for the 1st and 2nd Situations

In this section, the MSE of estimators T_1, T_2, T_3, $T_{\kappa_{min}}$ and of those presented in Appendix B.1 for the concrete case of two independent lognormal distributed variables will be estimated by means of simulations and compared for different parameter settings, such as μ and σ^2 and sample sizes n.

Shaban (1981) pointed out that the exact expressions of the MSE of his estimators are too complicated to permit analytical comparison between them. Therefore, for this distribution, a simulation study with 5000 repetitions, as performed in the last section, will be carried out for different sample sizes, i.e. $n = 10$, $n = 50$, $n = 100$ and $n = 1200$. The function to be estimated is: $\varrho = \exp[\mu_2]/\exp[\mu_1]$, see Equation B.1.3.

The simulations will be carried out under the following parameter settings:

- **PS 1:** $\mu_1 = \mu_2 = 1.1$ and $\sigma = 1/100$
- **PS 2:** $\mu_1 \neq \mu_2$, $\boldsymbol{\mu} = (0.1, 3)'$ and $\sigma = 1/100$
- **PS 3:** $\mu_1 \neq \mu_2$, $\boldsymbol{\mu} = (-1.40, 1.1)'$ and $\sigma = 1/100$. With $\exp[\mu_1] \approx 0$
- **PS 4:** $\mu_1 = \mu_2 = 1.1$ and $\sigma = 1/10$
- **PS 5:** $\mu_1 \neq \mu_2$, $\boldsymbol{\mu} = (0.1, 3)'$ and $\sigma = 1/10$
- **PS 6:** $\mu_1 \neq \mu_2$, $\boldsymbol{\mu} = (-1.40, 1.1)'$ and $\sigma = 1/10$. With $\exp[\mu_1] \approx 0$.

Simulation Results and Discussion for the 1st and 2nd Situations

In Table E.4 the estimated MSE of the estimators T_1, T_2, T_3 and $T_{\kappa_{min}}$ are compared with those of the estimators from literature, i.e. $\hat{\varrho}_{ML}$, $\hat{\varrho}_{MVUE_c}$, $\hat{\varrho}_{Shaban}$ and $\hat{\varrho}_{Shaban1}$, see Remark 6.5. As can be seen in the aforementioned table, the estimated MSE of all the presented estimators decreases when n increases. This decrease from $n = 10$ to $n = 1200$ is in most cases by two orders of magnitude, with the exception of T_2 and $\hat{\varrho}_{Shaban1}$ whose decreases are smaller for parameter settings 5 and 6.

When an evaluation is made on the basis of the parameter settings, it can be seen that all estimators have demonstrated to depend hardly on the parameter setting for μ and σ^2. They perform best when the means are equal and the standard deviation is small

(PS=1) and perform worst when the means are unequal, the standard deviation is large and additionally $\exp[\mu_1] \approx 0$ (PS=6).

The estimators proposed by Shaban (1981), denoted as $\hat{\varrho}_{Shaban}$ and $\hat{\varrho}_{Shaban1}$, declared as minimal MSE within the class B.1.5 and B.1.6 respectively, estimate the function $\varrho = \exp[\mu_2]/\exp[\mu_1]$, very accurately. However the estimator $T_{\kappa_{min}}$ was demonstrated to perform very well, even when it is compared with the aforementioned estimators.

It is difficult to state which estimator has the universally smallest MSE. In general all estimators, except T_2, can be recommended.

Assumed Parameter Settings (PS) for the 3rd Situation

These simulations have been made under the same parameter settings as described above. For the parameter settings 1-3 is $\sigma_1 = 1/100$ and $\sigma_2 = 1/10$ and for the parameter settings 4-6 is $\sigma_1 = 1/10$ and $\sigma_2 = 1$. The function to be estimated is: $\mathrm{E}[X_2]/\mathrm{E}[X_1] = \exp\left[(\mu_2 - \mu_1) + \frac{(\sigma_2^2 - \sigma_1^2)}{2}\right]$, see Equation B.1.1.

Simulation Results and Discussion for the 3rd Situation

In Table E.5 it can be seen that the estimated MSEs of all the presented estimators decrease when n increases. This decrease is from $n = 10$ to $n = 1200$ for all estimators, except for T_{ML}, by at least two orders of magnitude for all parameter settings.

T_{ML} performs well, for different parameter settings, but for parameter settings 5 and 6, i.e. larger standard deviation and $\mu_1 \neq \mu_2$, this estimator performs poorly and, as mentioned above, the decrease of its estimated MSE, by increasing the sample size, is of lower order of magnitude than with the rest of the estimators.

All estimators, except T_{ML} perform similarly, but for all parameter settings and sample sizes $T_{Shaban1}$ is even better than the estimator developed in this work $T_{\kappa_{min}}$.

In general all estimators, except T_2 and T_{ML} for parameter settings 5-6, can be recommended. But the estimator $T_{Shaban1}$ has the universally smallest estimated MSE.

Estimation of the RM of two Uncorrelated Gamma Variables

Suppose two independent random variables X_1 and X_2 distributed according to a gamma distribution with parameters α_1 and β_1 and α_2 and β_2 are given. For moments of the gamma distribution refer to Appendix A.1.4.

Now suppose that the ratio of means of the aforementioned random variables has to be estimated. This function is given as follows:

$$f(\boldsymbol{\mu}) = \frac{\mathrm{E}[X_2]}{\mathrm{E}[X_1]} = \frac{\alpha_2\beta_2}{\alpha_1\beta_1}.$$

The independent random sample means $\bar{x}_1 = \frac{1}{n}\sum_{i=1}^{n} X_{1i}$ and $\bar{x}_2 = \frac{1}{n}\sum_{i=1}^{n} X_{2i}$ are complete sufficient MVUEs of $\alpha_1\beta_1$ and $\alpha_2\beta_2$ respectively.

As mentioned in Section 2.1, Crow (1977) derived a Minimum Variance Unbiased Estimator (MVUE) from independent samples of the ratio of means of two gamma distributions of known shape.

This estimator is given by:

$$T_{CROW} = \frac{n\alpha_1 - 1}{n\alpha_1}\frac{\bar{x}_2}{\bar{x}_1}, \quad \text{with } n\alpha_1 > 1 \text{ and } \alpha_1 \text{ assumed to be known.}$$

It can be seen that with n sufficiently large $T_{CROW} \approx \frac{\bar{x}_2}{\bar{x}_1}$.

As following, the MSE of estimators T_1, T_2, T_3, $T_{\kappa_{min}}$ and T_{CROW} will be estimated by means of simulations and compared for different parameter settings, such as μ and σ^2 and sample sizes n.

The simulations will be carried out under the following parameter settings:

Assumed Parameter Settings

For this distribution a simulation study was performed for different sample sizes, i.e. $n = 10$, $n = 50$, $n = 100$ and $n = 1200$ and the following parameter settings:

- **PS 1: Equal Means and Variances (Both High)** ($\mu_1 = \mu_2$ and $\sigma_1^2 = \sigma_2^2$)

For $\boldsymbol{\alpha} = (4,4)'$ and $\boldsymbol{\beta} = (8,8)'$ is $\mu_1 = \mu_2 = 32$ and $\sigma_1^2 = \sigma_2^2 = 256$

- **PS 2: Equal Means and Variances (Both Medium)** ($\mu_1 = \mu_2$ and $\sigma_1^2 = \sigma_2^2$)

  For $\boldsymbol{\alpha} = (6,6)'$ and $\boldsymbol{\beta} = (2,2)'$ is $\mu_1 = \mu_2 = 12$ and $\sigma_1^2 = \sigma_2^2 = 24$

- **PS 3: Equal Means and Variances (Both Small)** ($\mu_1 = \mu_2$ and $\sigma_1^2 = \sigma_2^2$)

  For $\boldsymbol{\alpha} = (1/2,1/2)'$ and $\boldsymbol{\beta} = (1,1)'$ is $\mu_1 = \mu_2 = 0.5$ and $\sigma_1^2 = \sigma_2^2 = 0.5$

- **PS 4: Unequal Means and Variances (High μ_1)** ($\mu_1 \neq \mu_2$ and $\sigma_1^2 \neq \sigma_2^2$)

  For $\boldsymbol{\alpha} = (6,4)'$ and $\boldsymbol{\beta} = (6,4)'$ is $\mu_1 = 36$, $\mu_2 = 16$ and $\sigma_1^2 = 216$ and $\sigma_2^2 = 64$

- **PS 5: Unequal Means and Variances (Medium μ_1)** ($\mu_1 \neq \mu_2$ and $\sigma_1^2 \neq \sigma_2^2$)

  For $\boldsymbol{\alpha} = (4,2)'$ and $\boldsymbol{\beta} = (4,2)'$ is $\mu_1 = 16$, $\mu_2 = 4$ and $\sigma_1^2 = 64$ and $\sigma_2^2 = 8$

- **PS 6: Equal Means and Variances (Small μ_1)** ($\mu_1 \neq \mu_2$ and $\sigma_1^2 \neq \sigma_2^2$)

  For $\boldsymbol{\alpha} = (1/2,1)'$ and $\boldsymbol{\beta} = (1/2,1)'$ is $\mu_1 = 0.25$, $\mu_2 = 1$ and $\sigma_1^2 = 0.125$ and $\sigma_2^2 = 1$.

Simulation Results and Discussion

In Table E.6, the estimated MSE of a Minimum Variance Unbiased Estimator (MVUE), denoted by T_{CROW}, for the ratio of means of two gamma distributed random variables with **known shape**, is compared with the MSEs of the estimators T_1, T_2, T_3 and $T_{\kappa_{min}}$, as given in Remark 6.5.

It can be seen that, with the exception of T_2, the estimated MSEs of all the presented estimators decrease by at least two orders of magnitude as the sample size increases from $n = 10$ to $n = 1200$.

The estimated variance of the estimator T_2 tends to infinity, when μ_1 approaches zero, PS 4 and PS 5. T_3 is strongly influenced by estimator T_2 for μ_1 close to zero (remember that T_3 is a linear combination of estimators T_1 and T_2). The estimator T_{CROW} performs better than $T_{\kappa_{min}}$ for parameter settings 3 and 6, especially for small sample sizes.

In this case, it can be seen that for parameter settings 1,2,4 and 5 $T_{\kappa_{min}}$ has the smallest estimated MSE of all the estimators, especially for larger sample sizes.

6.3 Estimation of the Inverse of the Population Mean

The estimation of the inverse of the population mean and its functions, often arises in different sciences, such as in Physics and Biology. For instance, Allen (1957) described the use of the inverse of the mean, also known as the reciprocal mean, as a measure of covalent bond energy. This function is presented more formally in the following definition.

Definition 11 (Inverse of the Mean)

Let $\boldsymbol{x}$ be a random sample drawn from some distribution with mean μ and variance σ^2. The inverse of the mean *is defined as:*

$$f(\mu) = 1/\mu,$$

where $\mu \neq 0$.

Suppose $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)'$ represents a realisation of an i.i.d. random sample drawn from a probability distribution with mean μ and standard deviation σ.

Now, suppose that the inverse of the mean μ, see Definition 11, has to be estimated.

The general estimation approach presented in Section 2.2.3 will be applied for the statistical inference of the inverse of the mean. The general procedure has been summarised in Section 2.3. There it has been stated that the only condition for the application of the general estimation approach is the smoothness of the function to be estimated, i.e. the function has to be of at least class C^2, see Definition 1.

As follows it will be investigated, whether the function $f(\mu) = 1/\mu$ is of at least class C^2.

Remark 6.9 (Smoothness Condition)

Let $f(\mu) = 1/\mu$. The first and second order derivatives of $f(\mu)$ are given by:

$$f^{(1)}(\mu) = -1/\mu^2, \text{ and}$$

$$f^{(2)}(\mu) = 2/\mu^3.$$

It can be seen that the smoothness condition of function $f(\mu)$ is fulfilled.

For the estimation of $f(\mu) = 1/\mu$ the estimation approaches investigated in this work are given by:

$$T_1 = f(\hat{G}(\boldsymbol{x})) = 1/\bar{x} \qquad \text{and} \qquad T_2 = \hat{g}(f(x_i), i = 1, \cdots, n) = \tfrac{1}{n}\sum_{i=1}^{n} 1/x_i.$$

As second step of the estimation approach developed in this work, the generalised Jackknife approach, see Section A.1.3, is used in order to generate an unbiased approximated estimator for $f(\mu)$.

This estimator is given by:

$$T_3 = \frac{T_1 - RT_2}{1-R} = \frac{\left(1/\bar{x} - R\frac{1}{n}\sum_{i=1}^{n} 1/x_i\right)}{1-R}, \text{ with } R = \frac{\mathrm{b}(T_1, f(\mu))}{\mathrm{b}(T_2, f(\mu))}.$$

If an unbiased estimator is obtained, see Remark A.2, a linear adjustment as introduced by Troschke (2002), is applied in order to obtain a minimal Mean Squared Error (MSE) estimator $T_{\kappa_{min}}$. This estimator may be biased, but has smaller MSE than its unbiased counterpart generated by the generalised Jackknife approach.

For the calculation of the estimators T_3 and $T_{\kappa_{min}}$ the first and higher order moments of the estimators T_1 and T_2 are needed. They will be approximated by using the approach according to Neudecker and Trenkler (2005a) and enhanced in this work.

For the particular case of the estimation of $f(\mu) = 1/\mu$, an approximation on the basis of a linear plus quadratic function, of the function itself as well as of the estimators $T_1 = f(\bar{x})$ and $T_2 = \frac{1}{n}\sum_{i=1}^{n} f(x_i)$, as made in Remark 3.1, is presented as follows:

$$f^{Poly}(\mu) = a_0 + a\mu + A\mu^2,$$

$$T_1^{Poly} = f^{Poly}(\bar{x}) = a_0 + a\bar{x} + A\bar{x}^2, \text{ with } \bar{x} = \frac{1}{n}\mathbf{1}_n'\boldsymbol{x},$$

$$T_2^{Poly} = \frac{1}{n}\sum_{i=1}^{n} f^{Poly}(x_i) = a_0 + a\bar{x} + \frac{A}{n}\sum_{i=1}^{n} x_i^2.$$

where a_0, a and A are real constants. Furthermore, $a = f^{(1)}(\mu)$ and $A = \frac{1}{2}f^{(2)}(\mu)$.

Remark 6.10

Notice that replacing $a = f^{(1)}(\mu)$ *and* $A = \frac{1}{2}f^{(2)}(\mu)$ *by their respective expressions, given in Remark 6.14, the approximation* $f^{Poly}(\mu)$ *is equal to* a_0*. Suppose* $a_0 = f(\mu)$ *then* $f^{Poly}(\mu) = f(\mu)$*.*

In Section 3.1 it was pointed out that both, T_1^{Poly} and T_2^{Poly} are biased estimators for $f^{Poly}(\mu)$. The bias terms are presented in the following remark.

Remark 6.11 (Bias of Approximated Estimators T_1^{Poly} and T_2^{Poly})

The bias of estimators T_1^{Poly} *and* T_2^{Poly} *are given by:*

$$\mathrm{b}(T_1^{Poly}, f^{Poly}(\mu)) = A\sigma^2/n, \text{ and}$$

$$\mathrm{b}(T_2^{Poly}, f^{Poly}(\mu)) = A\sigma^2, \text{ see Definition 20.}$$

Furthermore, let $\alpha = A^2\sigma^4$*, then* $[\mathrm{b}(T_1^{Poly}, f^{Poly}(\mu))]^2 = \frac{\alpha}{n^2}$ *and* $[\mathrm{b}(T_2^{Poly}, f^{Poly}(\mu))]^2 = \alpha$ *are the square of the bias of estimators* T_1^{Poly} *and* T_2^{Poly}*, respectively.*

In the following remark the aforementioned estimators are presented in matricial notation.

Remark 6.12

For the estimation of $f^{Poly}(\mu)$ *the following estimators are proposed:*

$$\begin{aligned}
T_1^{Poly} &= f_0 + \boldsymbol{f}'\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{F}_1\boldsymbol{x},\\
T_2^{Poly} &= f_0 + \boldsymbol{f}'\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{F}_2\boldsymbol{x},\\
T_3^{Poly} &= f_0 + \boldsymbol{f}'\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{F}_3\boldsymbol{x},\\
T_{\kappa_{min}}^{Poly} &= \kappa_{min}(T_1^{Poly} - T_2^{Poly}) + T_2^{Poly},
\end{aligned}$$

with $\boldsymbol{x}$ *representing a realisation of an i.i.d. random sample and* f_0 *a real constant,* $\boldsymbol{f}$ *a* $n \times 1$ *vector and* $\boldsymbol{F}_l$*,* $l = 1, 2, 3$ *a* $n \times n$ *matrix.*

The aforementioned vector and matrices can be obtained by:

$$\boldsymbol{f} = \tfrac{a}{n}\mathbf{1}_n,\ \boldsymbol{F}_1 = \tfrac{A}{n^2}\mathbf{1}_n\mathbf{1}_n',\ \boldsymbol{F}_2 = \tfrac{A}{n}\boldsymbol{I}_n \text{ and } \boldsymbol{F}_3 = -\tfrac{A}{n(n-1)}\boldsymbol{I}_n - \mathbf{1}_n\mathbf{1}_n'.$$

6.3.1 Comparisons with Existing Solutions

For the estimation of the ratio of means or of the inverse mean Zellner (1978) has introduced an estimator that has shown to have (at least) finite first and second moments, and hence finite risk with respect to generalized quadratic loss. Additionally this estimator, known as the Minimum-Expected-Loss (MELO) estimator, is consistent, asymptotically efficient and asymptotically normal.

Zaman (1981) also investigated the estimation of the inverse of the mean of a normal distributed variable from a Bayesian point of view. Given a sufficiently large sample, Zaman (1981), Theorem 1, provides adequate justification for the ML estimator, if the loss function is bounded. He also studied some conditions under which the ML estimator may be more suitable than the MELO estimator, deduced by Zellner (1978), and vice versa, as well as situations in which neither is appropriate.

For definitions in the Bayesian approach, see Bernardo and Ramón (1998).

In this respect, Akahira and Takeuchi (1981) asserted that if a bounded loss function is appropriate, then lack of moments need not be regarded as a serious problem if the sample size is large enough.

Since, in this work, comparisons for finite sample sizes will be made, no more emphasis will be placed on the aforementioned estimators.

Based on the MELO estimator, Srivastava and Bhatnagar (1981) derived a class of estimators, which is free from the limitation of non-existence of moments. They derived exact expressions for the first two moments in the case of normal population and approximations for the non-normal case. These expressions are presented in Appendix B.2.1.

Voinov (1985) also derived unbiased estimators of powers of the inverse of population means, for the following cases:

a) unknown normally-distributed population mean μ and known variance σ^2 .

b) normal population mean $1/\mu^k, k = 1, 2, \ldots$, assuming μ and σ to be unknown. Additionally, $\mu > 0$ is assumed.

This approach is presented in Appendix B.2.2. In the same section it is also pointed out that the estimator deduced for the first case has infinite variance; and that for the second case an unbiased estimator of $1/\mu^k$ for $\mu < 0$ and σ^2 unknown does not exist, see Voinov (1985, p. 360) for discussions.

Since the estimator deduced by Voinov (1985) has **infinite** variance, it will not be possible to make a comparison with the estimators resulting from the estimation approach developed in this work, i.e. T_1, T_2, T_3 and $T_{\kappa_{min}}$.

The aforementioned estimators will be compared with those presented in Srivastava and Bhatnagar (1981), denoted by T_{Sriv} and presented in Appendix B.2.1, for different sample sizes, different underlying distributions and parameter settings in the following section.

Comparison of Estimators of the Inverse of the Mean

Most published methods concerning the estimation of the inverse of the mean are asymptotic in nature and based on the normal distribution. Since these assumptions are very restrictive, in this work the interest is also concentrated on comparing the performance of the estimators T_1, T_2, T_3, $T_{\kappa_{min}}$ and T_{Sriv} under different conditions, such as different sample sizes n, probability distribution assumptions and parameter settings. For the last condition it is especially interesting to observe how the estimators behave when the mean lies close to zero.

The simulation study to be carried out can be summarised as follows: The performance of the estimators T_1, T_2, T_3, $T_{\kappa_{min}}$ and T_{Sriv} will be compared under different probability distribution assumptions, parameter settings, such as μ and σ^2 and sample sizes n.

The approach developed by Srivastava and Bhatnagar (1981) is shown in Appendix B.2.1.

Assumed Distributions for the Estimation of the Inverse of the Mean

In order to make more generalisable comparisons between the estimators being analysed, different distributions will be simulated. These distributions are presented as follows.

Distribution I: Normal Distribution

Let X be normally distributed with mean μ and variance σ^2.

For this distribution Srivastava and Bhatnagar (1981) pointed out that the estimator T_{Sriv}, see Equation B.2.9, is asymptotically unbiased if $\mathcal{K} = 1$, while it has the smallest MSE to the order of their approximation, i.e. $(O(n^{-2}))$, if $\mathcal{K} = 4$. In this simulation study, $\mathcal{K} = 4$ will be used for the comparison of T_{Sriv} with the rest of the estimators.

Distribution II: Exponential Distribution

Let X be exponentially distributed with mean $\mu = \frac{1}{\lambda}$ and variance $\sigma^2 = \frac{1}{\lambda^2}$.

As mentioned above, Srivastava and Bhatnagar (1981) stated that the estimator T_{Sriv}, see Equation B.2.10, has smaller MSE than $1/\bar{x}$ for all negatively skewed populations and positively skewed populations with $\delta < 4$, provided $\mathcal{K}$ satisfies the inequality $0 < \mathcal{K} < 2(4 - \delta)$.

From literature it is known that the exponential distribution is a positively skewed distribution and that its Pearson's measure of skewness Sk is equal to 2, then it follows: $\delta = (\frac{2}{\mathfrak{v}})^{1/2}$, with $\mathfrak{v} = \frac{\sigma^2}{\mu^2}$.

Srivastava and Bhatnagar (1981) pointed out that for non-normal populations T_{Sriv} has a smaller MSE than $1/\bar{x}$ if $\mathcal{K} = (4 - \delta)$. For the comparison to be carried out in this work $\mathcal{K} = (4 - \delta)$ will be used.

Distribution III: Uniform Distribution

Let X be uniformly distributed in the interval $[a, b]$ with mean $\mu = \frac{a+b}{2}$ and variance $\sigma^2 = \frac{(b-a)^2}{12}$.

For this distribution the Pearson's measure of skewness Sk is equal to 0, then it follows that $\delta = 0$. As pointed out above if $\delta = 0$, then the expressions for the RB and RMSE are the same as those obtained for normal populations.

Assumed Parameter Settings for the Estimation of the Inverse Mean

Besides the three distribution assumptions presented above, for the estimation of $f(\mu)$, the performance of the estimators T_1, T_2, T_3, $T_{\kappa_{min}}$ and T_{Sriv} will be compared for different parameter settings of the random variable X.

The parameter settings for these distributions consist of a sequence of mean and variance values going from around zero to three.

Since T_{Sriv} has been deduced for large-sample approximations for non-normal distributions, in this simulation larger sample sizes than those used in the other simulations in this work will be used, so that it can be compared with the rest of the aforementioned estimators.

The performance of the aforementioned estimators will be compared for the sample sizes: $n = 250, 500, 750$ and 1200.

It is important to point out that the MSE of each estimator will be calculated using their approximated expressions $\text{MSE}(T_l^{Poly}, f^{Poly}(\mu))$, for $l = 1, 2, 3, \kappa_{min}, Sriv$ and estimated by simulations $\widehat{\text{MSE}}(T_l, f(\mu)) = \sum_{i=1}^{Rep} \frac{(T_{li} - f(\mu))^2}{Rep}$, for $l = 1, 2, 3, \kappa_{min}, Sriv$, with 5000 repetitions.

6.3.2 Results and Discussions from Approximated Expressions and from Simulations

Results of the approximated and simulated MSEs of different estimators of the inverse of the mean are presented in Figure F.1 - Figure F.14 of Appendix F.

In each figure two estimators are compared by means of: (a) the ratio of their approximated MSEs, i.e. $\frac{MSE(T_l^{Poly}, f^{Poly}(\mu))}{MSE(T_j^{Poly}, f^{Poly}(\mu))}$ for $l \neq j$, and (b) the ratio of their estimated MSEs, i.e. $\frac{\widehat{\text{MSE}}(T_l, f(\mu))}{\widehat{\text{MSE}}(T_j, f(\mu))}$ for $l \neq j$.

Two estimators are considered to perform similarly when the ratio of their MSEs is close to one.

For biparametrical distributions, e.g. the normal distribution, the figures are to be read as follows: On the y-axis the values of the ratios of the approximated or estimated MSEs between two estimators are presented and on the x-axis the values of a parameter, i.e. μ, of the distribution. The second parameter of this distribution, i.e. σ is represented in the figures with different colours.

For uniparametrical distributions (see next remark) the approximated or estimated MSEs of all estimators will be collocated together in the same figure. On the y-axis the values of the approximated or estimated MSEs are presented and on the x-axis the values of a parameter of the distribution. The different estimators are represented in the figures with different colours.

Remark 6.13

The mean and variance of the exponential distribution only depend on the parameter λ, while those of the uniform distribution depend on the parameters a and b.

Now, considering $b = a + constant$ the mean and the variance just depend on one parameter, i.e. $\mu = \frac{a+b}{2} = \frac{2a+constant}{2}$ and variance $\sigma^2 = \frac{(b-a)^2}{12} = \frac{(constant)^2}{12}$.

In this simulation study $constant = 2$ was used.

Distribution I. Normal distribution

In Figures F.1 and F.2 the performances of the estimators T_1 and $T_{\kappa_{min}}$ are compared by means of the ratio of their approximated and estimated MSEs.

It can be seen that the performances of the estimators T_1 and $T_{\kappa_{min}}$ are similar, especially for larger sample sizes. The ratios get closer to one as μ increases, especially for $\mu > 1$. For $\mu < 1$ the ratios vary depending on the standard deviation σ. The higher σ, the better the estimator $T_{\kappa_{min}}$ performs.

In Figures F.3 and F.4 the performances of the estimators T_1 and T_{Sriv} are also compared by means of the ratio of their approximated and estimated MSEs. It can be seen that for larger sample sizes the performance of the estimators T_1 and T_{Sriv} are similar, especially for $\mu > 1$. For $\mu < 1$ and smaller sample sizes, the ratios vary depending on the variance σ. The higher σ, the better the estimator T_{Sriv} performs and the more the ratios differ from one.

In Figures F.5 and F.6, it can be seen that the approximated and estimated MSEs of estimators T_1 and T_2 are different from one, i.e. the performances of these estimators differ considerably, especially for larger sample sizes. The ratios get closer to zero as μ increases, especially for $\mu > 1$, indicating that the estimator T_1 performs better than T_2.

In Figures F.7 and F.8, it can be seen that for small sample sizes and $\mu < 1$, the ratios between the approximated and estimated MSEs of $T_{\kappa_{min}}$ and T_{Sriv} vary depending on the standard deviation of the distribution, i.e. the higher the variability σ, the better the estimator T_{Sriv} performs. While for small σ and for $\mu > 1$ both estimators perform similarly and the ratios are close to one.

In general for the normal distribution, it can be seen that the ratios of the approximated and estimated MSEs of the presented estimators get closer to one, i.e. two estimators perform similarly, for increasing sample size, except T_2 which has not finite variance.

A pattern that could be observed for all comparisons under assumption of normal distribution is that the ratios differ for $\mu > 1$ and $\mu < 1$, being higher for $\mu < 1$. This can be explained by the fact that $f(\mu) = 1/\mu \approx \infty$, for $\mu \approx 0$.

Distribution II. Exponential distribution

In Figures F.9 and F.10, it can be seen that for the exponential distribution, with different values of the parameter λ, the approximated and estimated MSEs of all estimators, except of T_2 which is presented in Figure F.13, are almost identical, with decreasing MSE for increasing λ even for small sample sizes.

From this figures it can also be seen that $T_{\kappa_{min}}$ performs slightly better than the remaining estimators when λ is close to zero. However, T_{Sriv} performs better than the rest of the estimators for quite a few λ values between $(0, 1]$. Unfortunately, in this figure the small fluctuations along λ can hardly be appreciated.

Distribution III. Uniform distribution

In Figures F.11 and F.12, it can be seen that for the uniform distribution, with different values of the mean μ, the results are very similar to those obtained for the exponential distribution, with the difference that the approximated and estimated MSE values are slightly smaller. The performance of estimator T_2 is presented in Figure F.14

As a general conclusion it can be pointed out that for all distributions, $T_{\kappa_{min}}$ can be seen as a good alternative among the presented estimators.

It can also be seen that the approximated and estimated MSEs of this estimator are very close to that of T_{Sriv}. However, for the normal distribution when $\mu < 1$, T_{Sriv} has a smaller approximated and estimated MSE than that of the remaining estimators. On the other hand, for the exponential and uniform distributions, when λ and μ respectively are close to zero, the estimator $T_{\kappa_{min}}$ has a better performance.

The approximated MSE of estimator T_2 for the exponential and uniform distribution are presented in Appendix F.13 and F.14, respectively. From those figures it can be seen that the estimator T_2 cannot be recommended.

By comparing the approximated and estimated ratios of the MSEs of two estimators it can be observed that the approximated values, except the ratio between the MSEs of T_1 and $T_{\kappa_{min}}$ are smaller than those estimated for all distributions.

6.4 Estimation of the Odds in Favour of an Event

In probability theory and statistics, *the odds in favour of an event*, is the ratio of the probability of the occurrence of the event of interest to the probability that it does not occur. This statistical measure is commonly used in epidemiological studies to describe the likely harm an exposure might cause and is often estimated by the ratio of the number of times that the event of interest occurs to the number of times that it does not. This is presented more formally in the following definition.

Definition 12 (Odds in Favour of an Event.)

Assume that X is a Bernoulli distributed random variable with mean p. Then the odds in favour of an event *is defined as:*

$$f(p) = \frac{p}{1-p}, \text{ with } p \neq 0 \text{ and } p \neq 1.$$

If the odds of an event is greater than one, the event is more likely to happen than not, while the odds of an event that is certain to happen is infinite and the odds of an impossible event is zero.

The odds in favour of an event is very useful for the calculation of the *odds ratio*.

The *odds ratio* is defined as the ratio of the odds in favour of an event in one group to the odds of it occurring in another group. These groups might be men and women, an experimental group and a control group, or any other dichotomous classification. This concept is presented more formally in the following definition.

Definition 13 (Odds Ratio (OR))

Assume $P[A]$ and $P[B]$ are the probabilities of an event in two different groups A and B, respectively. Then the odds ratio *is given as:*

$$OR(A:B) = \frac{P[A](1-P[B])}{P[B](1-P[A])}.$$

The *odds ratio* is a measure useful for investigating whether the probability of a certain event is the same for two groups.

Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)'$ represent a realisation of an i.i.d. random sample of size n drawn from a Bernoulli probability distribution with mean p and variance $p(1-p)$. Now, suppose m successes are obtained from the random sample n.

The following function can be considered as an estimator for the parameter p:

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{m}{n}.$$

Now, suppose that the odds in favour of an event, see Definition 12, has to be estimated.

The general estimation approach presented in Section 2.2.3 will be applied for the statistical inference of the odds in favour of an event. The general procedure has been summarised in Section 2.3. There it has been stated that the only condition for the application of the general estimation approach is that of smoothness of the function to be estimated, i.e. the function has to be of at least class C^2, see Definition 1.

As follows it will be investigated, whether the function $f(p) = \frac{p}{1-p}$ is of at least class C^2.

Remark 6.14 (Smoothness Condition)

Let $f(p) = \dfrac{p}{1-p}$*. The first and second order derivatives of* $f(p)$ *are given by:*

$$f^{(1)}(p) = \frac{1}{(1-p)^2}, \text{ and}$$

$$f^{(2)}(p) = \frac{2}{(1-p)^3}.$$

It can be seen that the smoothness condition of function $f(p)$ is fulfilled.

For the estimation of the function $f(p)$ the estimation approaches investigated in this work are given by:

$$T_1 = f(\hat{G}(\boldsymbol{x})) = \frac{\hat{p}}{1-\hat{p}} = \frac{\frac{1}{n}\sum\limits_{i=1}^{n} x_i}{1-\frac{1}{n}\sum\limits_{i=1}^{n} x_i}, \text{ and}$$

$$T_2 = \hat{g}(f(x_i), i = 1, \cdots, n) = \frac{1}{n}\sum_{i=1}^{n} \frac{x_i}{1-x_i}.$$

Casella und Berger (1990, p. 292) have shown the invariance properties of the ML estimator. Thus, knowing that $\hat{p}$ is the ML estimator of p, it follows that $f(\hat{p}) = f(\hat{G}(\boldsymbol{x})) = \frac{\hat{p}}{1-\hat{p}}$ is the ML estimator of $f(p)$, see Definition 12.

Remark 6.15

Notice that as x_i, $i = 1, \cdots n$ are Bernoulli distributed, T_2 cannot be considered as an estimator since its denominator can take the value zero. It makes not possible the calculation of the estimators developed in this work.

The explanation of this problem and a possible solution are presented as follows.

In Definition 8 it is assumed that the function f is defined in $f : \mathfrak{B}^p \rightarrow \mathfrak{C}$, with p representing, in the multidimensional case, the number of variables and $\mathfrak{B}$, $\mathfrak{C} \subseteq \mathbb{R}$ the base set of the domain and counterdomain, respectively. For the univariate case, these function is defined in $f : \mathfrak{B} \rightarrow \mathfrak{C}$. Notice that the function $f(p) = \frac{p}{1-p}$ has as domain the set $(0, 1)$ and as counterdomain the set $(0, +\infty)$, i.e. $f : (0, 1) \rightarrow (0, +\infty)$. The same function is found in the estimator approaches presented above, i.e. $T_1 = f(\hat{p}) = \frac{\hat{p}}{1-\hat{p}}$ and $T_2 = \hat{g}(f(x_i), i = 1, \cdots, n)$. There it can be seen that the function $f(\hat{p})$ has the same base set in its domain and counterdomain as the function $f(p)$. On the other hand, the function $f(x_i), i = 1, \cdots, n$ has as domain the set $\{0, 1\}$ and as counterdomain the set $[0, +\infty]$, i.e. $f : \{0, 1\} \rightarrow [0, +\infty]$. As can be seen, in the estimator T_2 the function $f(x_i), i = 1, \cdots, n$ has not the same base sets in their domain and counterdomain as the functions $f(p)$ and $f(\bar{p})$ do, additionally they are not open intervals. Therefore, the estimation approach developed in this work cannot be applied for the estimation of the NLSS function $f(p) = \frac{p}{1-p}$.

Remark 6.16

In order to make it possible the calculation of the estimators developed in this work, the following approach is proposed:

1. *Generate k independent Bernoulli distributed random samples with parameter p, i.e. $(x_{11}, x_{12}, \cdots, x_{1n}), (x_{21}, x_{22}, \cdots, x_{2n}), \cdots, (x_{k1}, x_{k2}, \cdots, x_{kn})$.*

2. *Calculate the arithmetical mean in each random sample, i.e. $\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k$, with $\bar{x}_j \neq 1, \forall j$.*

3. *Calculate the estimators T_1 and T_2 for the function $f(p)$ from the sample sizes generated as above.*

For the estimation of the function $f(p)$ the estimators T_1 and T_2 are given by:

$$T_1 = f(\hat{G}(\bar{x})) = \frac{\bar{\bar{p}}}{1-\bar{\bar{p}}} = \frac{\frac{1}{k}\sum\limits_{j=1}^{k}\bar{x}_j}{1-\frac{1}{k}\sum\limits_{j=1}^{k}\bar{x}_j}, \text{ with } \bar{\bar{p}} = \frac{1}{k}\sum_{j=1}^{k}\bar{x}_j, \text{ and}$$

$$T_2 = \hat{g}(f(\bar{x}_j), j = 1, \cdots, k) = \frac{1}{k}\sum_{j=1}^{k}\frac{\bar{x}_j}{1-\bar{x}_j}, \text{ with } \bar{x}_j \neq 1, \forall j.$$

Now, the function $f(\bar{x}_j), j = 1, \cdots, k$ has the same base set in their domain and counter-domain as the functions $f(p)$ and $f(\bar{\bar{p}})$ do, i.e. $f : (0,1) \rightarrow (0, +\infty)$.

As second step of the estimation approach developed in this work, the generalised Jack-knife approach, see Section A.1.3, is used in order to generate an unbiased approximated estimator for $f(p)$.

This estimator is given by:

$$T_3 = \frac{T_1 - RT_2}{1-R} = \frac{\left(\frac{\frac{1}{k}\sum\limits_{j=1}^{k}\bar{x}_j}{1-\frac{1}{k}\sum\limits_{j=1}^{k}\bar{x}_j} - R\frac{1}{k}\sum\limits_{j=1}^{k}\frac{\bar{x}_j}{1-\bar{x}_j}\right)}{1-R}, \text{ with } R = \frac{\mathrm{b}(T_1, f(\mu))}{\mathrm{b}(T_2, f(\mu))}.$$

If an unbiased estimator is obtained, see Remark A.2, a linear adjustment as introduced by Troschke (2002), is applied in order to obtain a minimal Mean Squared Error (MSE) estimator $T_{\kappa_{min}}$. This estimator may be biased, but has smaller MSE than its unbiased counterpart generated by the generalised Jackknife approach.

For the calculation of the estimators T_3 and $T_{\kappa_{min}}$ the first and higher order moments of the estimators T_1 and T_2 are needed. They will be approximated by using the approach presented in Neudecker and Trenkler (2005a) and enhanced in the previous sections of this work.

For the particular case of the estimation of $f(p) = \frac{p}{1-p}$, an approximation on the basis of a linear plus quadratic function, of the function itself as well as of the estimators $T_1 = f(\bar{\bar{p}})$ and $T_2 = \frac{1}{k}\sum_{j=1}^{k} f(\bar{x}_j)$, as made in Remark 3.1, is presented as follows:

$$f^{Poly}(p) = a_0 + ap + Ap^2,$$

$$T_1^{Poly} = f^{Poly}(\bar{\bar{p}}) = a_0 + a\bar{\bar{p}} + A\bar{\bar{p}}, \text{ and}$$

$$T_2^{Poly} = \frac{1}{k}\sum_{j=1}^{k} f^{Poly}(\bar{x}_j) = a_0 + a\bar{\bar{p}} + \frac{A}{k}\sum_{j=1}^{k}\bar{x}_j^2,$$

where a_0, a and A are real constants. Furthermore, $a = f^{(1)}(p)$ and $A = \frac{1}{2}f^{(2)}(p)$.

Remark 6.17

Notice that replacing $a = f^{(1)}(p)$ and $A = \frac{1}{2}f^{(2)}(p)$ by their respective expressions, given in Remark 6.2, the approximation $f^{Poly}(p)$ is equal to $a_0 + \frac{p}{(1-p)^3}$.

In Section 3.1 it was pointed out that both, T_1^{Poly} and T_2^{Poly} are biased estimators for $f^{Poly}(p)$. The bias terms are presented in the following remark.

Remark 6.18 (Bias of Approximated Estimators T_1^{Poly} and T_2^{Poly})

Assume that X is a Bernoulli distributed variable with mean $\mu = p$ and variance

 $\sigma^2 = p(1-p)$. The bias of estimators T_1^{Poly} and T_2^{Poly} are given by:

$$b(T_1^{Poly}, f^{Poly}(p)) = A\sigma^2/n = \tfrac{p}{n(1-p)^2},\ \textit{and}$$

$$b(T_2^{Poly}, f^{Poly}(p)) = A\sigma^2 = \tfrac{p}{(1-p)^2},\ \textit{see Definition 20.}$$

In the following remark the aforementioned estimators are presented in matricial notation.

Remark 6.19

For the estimation of $f^{Poly}(p)$ the following estimators are proposed:

$$\begin{aligned} T_1^{Poly} &= f_0 + \boldsymbol{f}'\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{F}_1\boldsymbol{x}, \\ T_2^{Poly} &= f_0 + \boldsymbol{f}'\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{F}_2\boldsymbol{x}, \\ T_3^{Poly} &= f_0 + \boldsymbol{f}'\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{F}_3\boldsymbol{x}, \\ T_{\kappa_{min}}^{Poly} &= \kappa_{min}(T_1^{Poly} - T_2^{Poly}) + T_2^{Poly}, \end{aligned}$$

with $\boldsymbol{x}$ representing a realisation of an i.i.d. random sample and f_0 a real constant, $\boldsymbol{f}$ a $n \times 1$ vector and $\boldsymbol{F}_l$, $l = 1, 2, 3$ a $n \times n$ matrix.

The aforementioned vector an matrices can be obtained by:

$$\boldsymbol{f} = \tfrac{a}{n}\mathbf{1}_n,\ \boldsymbol{F}_1 = \tfrac{A}{n^2}\mathbf{1}_n\mathbf{1}_n',\ \boldsymbol{F}_2 = \tfrac{A}{n}\boldsymbol{I}_n \text{ and } \boldsymbol{F}_3 = -\tfrac{A}{n(n-1)}\boldsymbol{I}_n - \mathbf{1}_n\mathbf{1}_n'.$$

The approximated and estimated MSE of the estimators presented in Remark 6.19 will be compared under assumption of the following parameter setting.

Assumed Parameter Settings (PS)

For a comparison of the estimators presented in Remark 6.19 the approach outlined in Remark 6.16 will be used.

In this case $k = 4$ independent Bernoulli distributed random samples with mean p and variance $p(1-p)$ are generated, i.e. $(x_{11}, x_{12}, \cdots, x_{1n})$, $(x_{21}, x_{22}, \cdots, x_{2n})$, $(x_{31}, x_{32}, \cdots, x_{3n})$ and $(x_{41}, x_{42}, \cdots, x_{4n})$, with $n = 250, 500, 750$ and 1200. As second step the arithmetical mean of each random sample, i.e. $\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4$ is obtained. Then the estimators T_1,T_2, T_3

and κ_{min} for the estimation of the function $f(p)$ are calculated. Finally, the approximated and estimated MSEs of the aforementioned estimators are calculated.

Assume that X is a Bernoulli distributed variable with mean p and variance $p(1-p)$.

Parameter Setting (PS) 1. $p = 0.02$, $1 - p = 0.98$.

$\mathrm{E}[X] = p = 0.02$

$\mathrm{var}(X) = p(1-p) = 0.0196$

Φ=$\mathrm{E}[Z^3] = \mathrm{E}[X-\mu]^3 = p(1-p)(1-2p) = 0.018816$

Ψ=$\mathrm{E}[Z^4] = \mathrm{E}[X-\mu]^4 = p(1-p)(3p^2-3p+1) = 0.018448$

Parameter Setting (PS) 2. $p = 0.1$, $1 - p = 0.9$.

$\mathrm{E}[X] = p = 0.15$

$\mathrm{var}(X) = p(1-p) = 0.09$

Φ=$\mathrm{E}[Z^3] = 0.072$

Ψ=$\mathrm{E}[Z^4] = 0.0657$

Parameter Setting (PS) 3. $p = 0.5$, $1 - p = 0.5$.

$\mathrm{E}[X] = p = 0.5$

$\mathrm{var}(X) = p(1-p) = 0.25$

Φ=$\mathrm{E}[Z^3] = 0$

Ψ=$\mathrm{E}[Z^4] = 0.0625$

Parameter Setting (PS) 4. $p = 0.9$, $1 - p = 0.1$.

$\mathrm{E}[X] = p = 0.9$

$\mathrm{var}(X) = p(1-p) = 0.09$

Φ=$\mathrm{E}[Z^3] = -0.072$

Ψ=$\mathrm{E}[Z^4] = 0.0657$

6.4.1 Results and Discussions from Approximated Expressions and from Simulations

Numerical results of the approximated MSEs or $MSE(T_l^{Poly}, f^{Poly}(\mu))$, $l = 1, 2, 3, \kappa_{min}$ and of the estimated MSE or $\widehat{\text{MSE}}(T_l, f(\mu))$, $l = 1, 2, 3, \kappa_{min}$ for different sample sizes n and parameter settings, presented above, are shown in Appendix G by means of a table. This table contains the approximated and estimated MSE of the aforementioned estimators on the left and right side, respectively. Each estimator is presented in a column. The 4 different parameter settings (PS), as given above, were accommodated as block rows, where each block row consists of 4 rows representing 4 different sample sizes. In this way, the MSE of each estimator for a given parameter setting and a given sample size can be compared simultaneously. This accommodation on the table makes it possible to compare, for example, the approximated and estimated MSEs of each estimator across the different sample sizes by moving 1 position down in each block row.

An illustration by means of an arbitrary example is presented as follows:

Consider the value 3.345e-04 in row number 6 of the right side table in Table G.1, i.e. second row in the second block row and in the first column. This value shows the approximated MSE of estimator T_1 calculated from 4 Bernoulli distributed samples, with p=0.1 and $1 - p$=0.9, each one of them consisting of 500 elements.

Moving 1 top-down position, but staying in the same block row, it can be seen that the approximated MSE of the same estimator takes the value 2.568e-04 when the samples, as described above, consist of 750 elements instead of 500 and that moving one row lower, the approximated MSE of the same estimator takes the value 1.491e-04 when the samples consist of 1200 elements. From this comparison among the sample sizes, it can be seen that the approximated MSE of estimator decreases as the sample size increases.

Now, keeping the same initial value 3.345e-04 and moving 4 top-down positions, i.e. moving to the second row of the second block row, it can be seen that the approximated MSE of the same estimator takes the value 7.755e-02 when the Bernoulli distributed samples, also consisting of 500 elements, have the parameters p=0.5 and $1 - p$=0.5. This indicates a notable increase in the approximated MSE of estimator T_1, when the parameter p changes from 0.1 to 0.5.

The main findings from the results presented in the Table G.1 can be compared by keeping one or two parameters of the simulation fixed and varying the rest.

- Keeping the PS fixed and varying n it can be seen that for all PS the approximated and estimated MSE of all estimators decreases, in much cases by one order of magnitude, as the sample size increases.
- Keeping n fixed and varying PS it can be seen that the estimated MSEs of all estimators for PS 3 and 4 are bigger, while no meaningful differences among the first two PS are observed, whereas the smallest MSEs are obtained with PS 1 and the largest with PS 4, where p is closer to 1, it has already been mentioned that for $p \approx 1$ then $\frac{p}{1-p} \approx \infty$.
- Keeping n and PS fixed and varying T_l, for $l = 1, 2, 3, \kappa_{min}$ it can be seen that the approximated and estimated MSE of the estimator $T_{\kappa_{min}}$ is smaller than those of the remaining estimators for all parameter settings and sample sizes, whereas for PS 3 estimators T_3 and $T_{\kappa_{min}}$ perform similarly.

From the exposed above, it can be seen that for the estimation of $f(p) = \frac{p}{1-p}$, the smaller the probability p, the smaller the MSEs of all estimators. That means:

$\widehat{\text{MSE}}(T_l)_{PS1} < \widehat{\text{MSE}}(T_l)_{PS2} < \widehat{\text{MSE}}(T_l)_{PS3} < \widehat{\text{MSE}}(T_l)_{PS3}$, $l = 1, 2, 3, \kappa_{min}$, and

$\text{MSE}(T_l^{Poly})_{PS1} < \text{MSE}(T_l^{Poly})_{PS2} < \text{MSE}(T_l^{Poly})_{PS3} < \text{MSE}(T_l^{Poly})_{PS4}$, $l = 1, 2, 3, \kappa_{min}$.

A peculiarity presented in the estimation of this particular function is that the approximated MSEs are higher than the estimated MSEs. This can be explained by the use of the estimation approach presented in Remark 6.16.

Chapter 7

Summary, Conclusions and Plans for Further Research

Contents

In this work a new approach to the statistical inference of nonlinear, sufficiently smooth (NLSS) functions has been introduced. The approach proposed here is based on Taylor expansions to approximate sufficiently smooth functions and on the development of a minimal Mean Square Error (MSE) estimator of this approximation. This procedure has been explored very carefully, from mathematical and application-oriented points of view.

The work-flow of this study as well as its most significant contributions are summarised as follows.

7.1 Summary and Conclusions

Chapter 2 of this study was devoted to an in-depth review of literature on the estimation of NLSS functions. Apart from framing the research's focus, providing readers with an overview of, and background to the most significant definitions, results and progresses in this area, the literature review chapter served to direct the research towards an in-depth exploration of comparatively unexplored issues without solutions that have been controversial and that have prompted several discussions. This brings us directly to the question of the research contribution of this study. The research has made different contributions to the field of point estimation of NLSS functions, some of which shall now be briefly highlighted.

The first contribution concerns the recognition of the restrictions that existing estimation methods present. The literature revealed that most published methods concerning the estimation of NLSS functions are asymptotic in nature and based on the assumption of normal distribution, and that the random variables involved are not correlated. Experience confirms that this is not always the case, i.e. the conditions under which those methods have been generated are very restrictive, so that the generated results are based upon the particular case and are therefore only representative for that case. Additional arguments are therefore required in order to generalise the results, see for example Qiao et al. (2006), who pointed out that the development of a satisfactory estimator of the ratio, when the involved random variables are dependent, remains an area for future research.

It could also be argued that many of the studies discussed in Chapter 2 are incomplete or meager. Many research projects have uncovered a variety of conclusions regarding the performance of their estimation methods and researchers have benefited from the properties of the normal distribution. For example, for the estimation of the ratio of means when the variables are obtained from two independent distributions, the estimation of such a function presents different problems, for instance the aforementioned methods do not perform well in a range of small-sampling settings such as a small-sample bio-availability study.

The second contribution draws directly from the first. From an exhaustive literature research it could be stated that several studies show conflicting results, which indicates the need for further research. In this respect, as an example, Srivastava and Bhatnagar (1981) have erroneously stated that the Maximum Likelihood estimator of the inverse of the mean has no finite moments. Conversely, Voinov (1985) has demonstrated that this statement holds only for the second and higher order moments.

A third piece of research contribution lies in the acquisition of knowledge about two useful estimation approaches, as presented in Definition 8. These estimation approaches are the building blocks for the construction of a new inference method of NLSS functions developed in this study. For the particular case of the estimation of the ratio of means, Rao (2002) called them "mean of ratios" and "ratio of means". The general procedure of the new inference method was also presented in Chapter 2. This inference method is basically motivated by Casella and Berger (1990), Neudecker and Trenkler (2002) and Troschke (2002). Each step of this general procedure is illustrated through the work-flow of this dissertation as follows:

The first step consist of approximating the NLSS function and estimating it by using the aforementioned estimation approaches, based on the approach presented by Casella and Berger (1990). This was worked out in Chapters 2 and 3, where different basic properties, such as the means and variances, of the aforementioned estimation approaches were described and approximated expressions were deduced, following the approach presented by Neudecker and Trenkler (2005a). For both multivariate and univariate data, the normal distribution was considered as a special case. Based on the request presented in Qiao et al. (2006) for the multivariate distributions, the properties of estimators for correlated random variables were also considered as a special case and, in the same chapter, a new approach based on Kleffe and Rao (1988, Section 2.1) was developed.

For the second step of the general procedure of the new inference method, the generalised Jackknife approach was used in Chapter 4 in order to generate an unbiased estimator for the approximation of NLSS functions.

After the unbiased estimator was obtained, a linear adjustment as presented in Troschke (2002) was applied as the third step of the general procedure of the new inference method, in order to obtain a minimal Mean Squared Error (MSE) estimator. This was presented in Chapter 5.

Besides the theoretical results presented in Chapters 2 - 5 which give a general framework, in Chapter 6, application-specific estimations have been developed using the results from those chapters. Asymptotical results for the estimation of NLSS from various literature studies were compared with both the new estimator, developed using the generalised Jackknife approach and the estimator obtained using the approach introduced by Troschke (2002).

As the first NLSS function, the ratio of means was considered in this chapter. Additionally, the estimation approaches for the ratio of means of two lognormal and gamma distributed variables were also presented as special cases.

Estimations of the inverse of the mean and of the odds in favour of an event were also considered in Chapter 6 as examples of a NLSS function for which the estimation approach deduced in this work can be applied. In the same chapter, comparisons between estimators presented in literature and those analysed and developed in this work were also carried out.

The new inference method is not based on any assumption of any type of distribution or of any data structure. In this work, different simulations were carried out in order to observe how the new method works under different distribution assumptions and sample sizes. This method can even be used for estimating the parameters of two correlated distributions.

For the comparisons between different estimators, carried out by means of simulations, in Chapter 6, it was necessary to introduce some clarifying aspects and settings which are fundamental for those simulations. Thus, a method for generating two correlated random variables was deduced, so that the comparisons between the estimators can be carried out under the assumption of different (user-defined) correlation levels and distributions.

The performance of the aforementioned estimators was compared by means of their approximated and simulated MSEs under different distribution assumptions, parameter settings, correlation coefficients and sample sizes. As a performance measure, the MSE was used.

The results of the comparisons suggest that the estimators developed here were very convincing in nearly all the situations presented. They compared favorably to existing standard methods. In general, some of the results obtained from the simulation study can be summarised as follows:

- The estimator T_2 is the worst estimator. This conclusion agrees with the properties exposed in Rao (1952), where the role of the estimators T_1 and T_2 based on pairs of observations from normal populations, was discussed. T_2 was shown to be **inconsistent** while T_1 was considered **consistent**.
- T_3 can be recommended for large sample sizes. This estimator has notably been improved for most distributions and parameter settings through the use of Lemma 5.1, that is $\text{MSE}(T_{\kappa_{min}}) < \text{MSE}(T_3)$.
- A final observation is that the results of this study provide support for the new estimator, since it is demonstrated to have a minimal Mean Squared Error even when it was compared with Minimal Variance Unbiased Estimators.

The new approach developed here has improved the existing theoretical and practical results for the estimation of NLSS functions of distribution parameters given by its minimal MSE.

7.2 Further Plans for Research

By virtue of this being a new approach of estimating NLSS functions, a series of interesting and challenging problems is brought up - both from a purely mathematical viewpoint, as well as from the perspective of applications. Although those problems raised could be addressed by taking solely one viewpoint or the other, according to one's experience, the best solutions can be obtained by means of a combination of mathematical theory and supporting simulation-based, application-oriented evaluation.

To some extent this has been achieved in this work, wherein a balance between the asymptotic theory and real-life, small-sample approximations has been maintained.

Throughout this work, wherever appropriate, there have been suggestions for modification, improvement and new proposals for further research. There are also many more possible suggestions and unanswered questions, which are addressed as follows:

- Other NLSS functions of distribution parameters (e.g. the ratio of two variances) should be explored in order to establish whether the estimator deduced in this work is generally the best choice. Nikulin and Voinov (1993 and 1996) have presented different tables with unbiased estimators for different functions of different parameters for the most commonly used distributions. It would be interesting to compare the estimator deduced here with the unbiased estimator for some functions presented in these works, for which the variance is not infinite. There is also a variety of statistical problems, which can be viewed as questions of inference on NLSS functions of the parameters in the general linear regression model. Inference for the general formulation of this problem has only been developed using the Bayesian approach, and credibility intervals for individual functions as well as for linear combinations of the functions of the parameters have been obtained by using numerical integration.

- The estimators developed in this work should be compared with examples taken from the bibliography, where the normal distribution is less common than other shifted distributions (e.g. the two parameter or shifted exponential distribution, shifted Gamma, shifted Weibull, Lognormal, etc.), Meeker and Hahn (1980) considered the case where the involved random variables follow an exponential time-to-failure distribution.

- To go further than the point estimation for $f(\boldsymbol{\mu})$ and investigate how approximated confidence intervals for such a function could be determined, the point estimation obtained in this work should be used. From a frequentist point of view, the pointwise estimation of NLSS functions can be made using the approach presented in this study. It is well known, however, that several methodological difficulties arise if the objective is to calculate interval estimates for this parameter, see Fieller and Creasy (1954). An interesting approach would be to use the point estimation obtained in this work to calculate approximated confidence intervals for NLSS functions of distribution parameters for both normal and non-normal distribution, where the sample size is not necessarily large.

Chapter 8

References

Akahira, M. and Takeuchi, K. (1981). *Asymptotic efficiency of statistical estimators: Concepts and higher order asymptotic efficiency (Lecture Notes in Statistics 7).* Springer-Verlag, 1981.

Allen, T. (1957). *Reciprocal mean as a measure of covalent bond energy.* The Journal of Chemical Physics, 27, 810.

Altman, D. and Bland, M. (2000). *Statistics note: The odds ratio.* British Medical Journal, 320, 1468.

Barndorff-Nielsen, O. (1983). *On a formula for the distribution of the maximum likelihood estimator.* Biometrika, 70, 343 - 365.

Bernardo, J. (1977). *Inference about the ratio of normal means: A Bayesian approach to the Fieller-Creasy problem.* Recent Developments in Statistics. Amsterdam: North-Holland, 345 - 350.

Bernardo, J. and Ramón, J. (1998). *An introduction to Bayesian reference analysis: Inference on the ratio of multinomial parameters.* The Statistician 47, 1 - 35.

Beyene, J. and Moineddin, R. (2005). *Methods for confidence interval estimation of a ratio parameter with application to location quotients.* Medical Research Methodology. Online at http://www.biomedcentral.com/content/pdf/1471-2288-5-32.pdf.

Buonaccorsi, J. (1985). *Ratios of linear combinations in the general linear model.* Communications in Statistics. Theory and Methods, 14(3), 635 - 650.

Casella, G. and Berger, R. (1990). *Statistical inference.* Wadsworth and Brooks/Cole, Pacific Grove.

Casella, G. and Khuri, A. (2002). *The existence of the first negative moment revisited.* The American Statistician, 56(1), 44 - 47.

Casella, G. and Piegorsch, W. (1985). *The existence of the first negative moment.* The American Statistician, 29(1), 60 - 62.

Creasy, M. (1954). *Limits for the ratio of the means.* Journal Royal Statistical Society, Serie B(16), 186 - 194.

Crow, E. (1977). *MVUE of the ratio of means of two lognormal variates and two Gamma variates.* Communications in Statistics. Theory and Methods, A6(10), 967 - 975.

Fieller, E. (1932). *Distribution of the index in a normal bivariate population.* Biometrika, 24, 428 - 440.

Fieller, E. (1944). *A fundamental formula in the statistics of biological assay and some applications.* Quarterly Journal of Pharmacy and Pharmacology. 17, 117 - 123.

Fieller, E. (1954). *Some problems in interval estimation.* Journal of the Royal Statistical Society, Serie B(16), l75 - l85.

Finney, D. (1971). *Probit analysis.* London: Cambridge University Press.

Fitzgerald, S., et al. (2006). *Generating correlated matrix exponential random variables.* Advances in Engineering Software, 37(2).

Förster, S. (1997). *Monte Carlo simulation of correlated random variables.* http://citeseer.ist.psu.edu/432929.html.

Frauendorf, E. and Trenkler, G. (1998). *On the exchangeability of transformations and the arithmetic mean.* Journal of Statistical Computation and of the Simulation, 61, 305 - 307.

Frauendorf, E., Neudecker, H. and Trenkler, G. (2005). *Multivariate data, the arithmetic mean and exchangeability of transformation.* Linear Algebra and its Applications 410, 87 - 95.

Friedrich, J., et al. (2008). *The ratio of means method as an alternative to mean differences for analyzing continuous outcome variables in meta-analysis: A simulation study.* Medical Research Methodology. Online at http://www.biomedcentral.com/1471-2288/8/32.

Gleser, L. and Hwang, J. (1987). *The nonexistence of* $100(1-\alpha)\%$ *confidence sets of finite expected diameter in errors-in-variables and related models.* Annals of Statistics, 15, 1351 - 1362.

Gnot, S., Trenkler, G. and Zmyślony, R. (1995). *Nonnegative minimum biased quadratic estimation in the Linear Regression Models.* Journal of Multivariate Analysis, 54, 113 - 125.

Gradshteyn, I. S. and Ryzhik, I. M. (2000). *Hessian determinants.* Tables of Integrals, Series, and Products, 6th ed. San Diego, CA: Academic Press, 1069.

Gray, H. and Schucany, W. (1972). *The generalized Jackknife statistic.* New York: Marcel Dekker.

Hauschke, D., et al. (1999a). *Sample size determination for proving equivalence based on the ratio of two means for normally distributed data.* Statistics in Medicine, 18, 93 - 105.

Hauschke, D., et al. (1999b). *Proof of safety in toxicology based on the ratio of two means for normally distributed data.* Biometrical Journal, 41, 3, 295 - 304.

Hinkley, D.(1969). *On the ratio of two correlated normal random variables.* Biometrika, 56(3), 635 - 639.

Kleffe, J. and Rao, C. (1988). *Estimation of variance components and applications.* North Holland.

Kleijnen, J. (1974). *A note on sampling two correlated variables.* Simulation, 22(2), 45-46.

Lehmann, E. (1983). *Theory of point estimation.* Wiley, New York.

Lehmann, E. (1996). *Testing statistical hypotheses.* 2nd Edition, Wiley, New York.

Lehmann, E. and Shaffer, J. (1988). *Inverted distributions.* The American Statistician, 42, 191 - 194.

Magnus, J. and Neudecker, H. (1979). *The commutation matrix: Some properties and applications.* The Annals of Statistics, 7, 381 - 394.

Malley, J. (1982). *Simultaneous confidence intervals for ratios of normal means.* Journal of the American Statistical Association, 77, 170 - 176.

Meeker, W. and Hahn, G. (1980). *Prediction intervals for the ratios of normal distribution sample variances and exponential distribution sample means.* Technometrics, 22(3), 357 - 366.

Mendenhall, W. and Lehman, E. (1960). *An approximation to the negative moments of the positive binomial useful in life testing.* Technometrics 2, 227 - 242.

Mood, A., Graybill, F. and Boes, D. (1974). *Introduction to the theory of statistics.* 3rth Edition, McGraw-Hill.

Neudecker, H. and Trenkler, G. (2002). *Third and fourth moment matrices of Vec$\boldsymbol{X}'$ in multivariate analysis.* Linear Algebra and Its Applications, 354, 223 - 229.

Neudecker, H. and Trenkler, G. (2005a). *On the Approximate Variance of a Nonlinear Function of Random Variables.* Contributions to Probability and Statistics: Applications and Challenges. Proceedings of the International Statistics Workshop. University of Canberra, 172 - 177.

Neudecker, H. and Trenkler, G. (2005b). *Estimation of the Kronecker and inner product of two mean vectors in multivariate analysis.* Discussiones Mathematicae - Probability and Statistics. 25(2), 207 - 215.

Pandey, B. and Malik, H. (1990). *Estimation of the mean and the reciprocal of the mean of the inverse Gaussian distribution.* Communications in Statistics - Theory and Methods, 18, 1187 - 1201.

Qiao, C., et al. (2006). *Comparison of two common estimators of the ratio of the means of independent normal variables in agricultural research.* Journal of Applied Mathematics and Decision Sciences. Online at www.emis.de/journals/HOA/JAMDS/0d5a.pdf.

Quenouille, M. (1956). *Notes on bias in estimation.* Biometrika, 43(3/4), 353 - 360.

Rao, T. (2002). *Mean of ratios or ratio of means or both?.* Journal of Statistic Planning and Inference, 102, 129 - 138.

Shorack, G. (2000). *Probability for statisticians.* Springer, New York.

Schmidt, K. and Trenkler, G. (2006). *Einführung in die moderne matrix-algebra: Mit Anwendungen in der Statistik.* Berlin et al., Springer.

Schucany, W., et al. (1971). *On bias reduction in estimation.* Journal of the American Statistical Association, 66(335), 524 - 533.

Seber, G. (1977). *Linear regression analysis.* John Wiley, New York.

Shaban, S. (1981). *On the estimation of the ratio of means and other characteristics of two lognormal variates.* Biometrical Journal, 23, 357 - 369.

Shilov, G. and Gurevich, B., (1978). *Integral, measure, and derivative: A Unified Approach.* Dover Publications.

Srivastava, V. and Bhatnagar, S. (1981). *Estimation of the inverse of mean.* Journal Statistical Planning and Inference, 5, 324 - 334.

Tiwari H. and Elston R. (1999). *The approximate variance of a function of random variables.* Biometrical Journal 41, 351 - 357.

Troschke, S. (2002). *Enhanced approaches of the combination of forecasts.* Josef Eul Verlag.

Tukey, J. (1958). *Bias and confidence in not quite large samples.* Annals of Mathematical Statistics, 29, 614.

Van der Merwe, A. (1992). *Bayesian inferences on nonlinear functions of the parameters in linear regression.* Annals of the Institute of Statistical Mathematics, 44(2), 201 - 211.

Voinov, V. (1985). *Unbiased estimators of powers of the inverse of mean and related problems.* Sankhyā: The Indian Journal of Statistics, 47(B), 354 - 364.

Voinov, V. and Nikulin, M. (1993). *Unbiased estimators and their applications, Vol. 1: Univariate Case.* Kluwer Academic Publishers.

Voinov, V. and Nikulin, M. (1996). *Unbiased estimators and their applications, Vol. 2: Multivariate Case.* Kluwer Academic Publishers.

Whittaker, E. and Watson, G. (1990). *A Course in modern analysis.* Cambridge, England: Cambridge University Press.

Wu, J., et al. (2002). *Likelihood analysis for the ratio of means of two independent lognormal distributions.* Biometrics, 58, 463 - 469.

Zacks, S. (1971). *The theory of statistical inference.* New York: Wiley.

Zaman, A. (1981). *Estimators without moments: The case of the reciprocal of a normal mean.* Journal of Econometrics, 15(2), 289 - 298.

Zellner, A. (1978). *Estimator of functions of population means and regression coefficients including structural coefficients: A Minimum Expected Loss (MELO) approach.* Journal of Econometrics 8, 127 - 158.

Zohu, X., et al. (1997). *Methods for comparing the means of two independent lognormal samples.* Biometrics 53, 1129 - 1135.

Part I

Appendixes

Appendix A

Additional Definitions, Remarks and Lemmas

A.1 General Definitions

A.1.1 Functions

This section commences with a review of the definition of a *function*. This can be considered as a complement of Section 2.2.1.

Definition 14 (Function)

A function, *say* f, *with domain* $\mathfrak{B}$ *and counterdomain* $\mathfrak{C}$, *where* $\mathfrak{B}$ *and* $\mathfrak{C}$ *are unions of open intervals, i.e.* $f : \mathfrak{B} \to \mathfrak{C}$, *is a collection of ordered pairs, say* $(\mathfrak{b}, \mathfrak{c})$, *satisfying* (i) $\mathfrak{b} \in \mathfrak{B}$ *and* $\mathfrak{c} \in \mathfrak{C}$; (ii) *each* $\mathfrak{b} \in \mathfrak{B}$ *occurs as the first element of some ordered pair in the collection (each* $\mathfrak{c} \in \mathfrak{C}$ *is not necessarily the second element of some ordered pair); and* (iii) *no two (distinct) ordered pairs in the collection have the same first element.*

The set of all values of f is called the *range* of f, i.e. $f = \{\mathfrak{c} \in \mathfrak{C} : \mathfrak{c} = f(\mathfrak{b}) \text{ for some } \mathfrak{b} \in \mathfrak{B}\}$ and is always a subset of the counterdomain $\mathfrak{C}$ but is not necessarily equal to it.

Definition 15 (Linear Functions)

Consider a function f, *say* $f : \mathfrak{B} \to \mathfrak{C}$. *Linear functions are functions that have the form:*

$$f(x) = mx + n \; ; \quad m, n \in \mathbb{R}.$$

In a graph m is the slope of the line $y = mx + n$ and n is the y-intercept. A linear function has a constant slope and is said to be increasing (or rising) when $m > 0$ and is decreasing (or falling) when $m < 0$. The graph of a linear function is a straight line. In the special case where $m = 0$ the function $f(x) = n$ is a constant function.

Functions whose graphs are not straight lines are called *nonlinear functions*. The graph of a nonlinear function can be a curved line, whose slope changes for at least one x. More formally, this kind of functions is introduced in the following definition.

Definition 16 (Nonlinear Functions)

Nonlinear functions are all functions other than linear ones.

Notice that although the slope of a linear function m is the same no matter where on the line it is measured, the slope of a nonlinear function can be different at each point on the line. Thus, there is no constant slope for a nonlinear function.

In several fields of statistics, linear functions are used to explain the relationship between variables. Nevertheless, there are also different fields of statistics, such as econometrics, where a linear function cannot explain the relationship between variables. In such cases a nonlinear function tends to be more appropriate.

At this point it is important to introduce a measure of how a function $f(x)$ changes as its input x changes. This measure is known as *derivative* of $f(x)$ with respect to x.

Definition 17 (Derivatives of a Function)

The simple derivative *of a function f with respect to the variable x is denoted by either $f^{(1)}(x)$ or $\frac{d^{(1)}f(x)}{dx}$ and defined as:*

$$f^{(1)}(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}.$$

If the limit exists, then f is differentiable *at x. The* i-the derivative *is denoted by either $f^{(i)}(x)$ or $\frac{d^{(i)}f(x)}{dx^i}$.*
When a function $f(x_1, x_2, \cdots, x_p)$ depends on more than one variable, a partial derivative $\frac{\partial f(x_1,x_2,\cdots,x_p)}{\partial x_1}, \cdots, \frac{\partial f(x_1,x_2,\cdots,x_p)}{\partial x_p}, \frac{\partial^{(2)} f(x_1,x_2,\cdots,x_p)}{\partial x_1 x_2}, \cdots, \frac{\partial^{(2)} f(x_1,x_2,\cdots,x_p)}{\partial x_1 x_p}, \cdots$ can be used to specify the derivative with respect to one or more variables, with:

$$\frac{\partial f(x_1, x_2, \cdots, x_p)}{\partial x_m} = \lim_{h \to 0} \frac{f(x_1, ..., x_m + h, ..., x_p) - f(x_1, ..., x_m, ..., x_p)}{h}.$$

If x and $f(x)$ are real numbers, and if the graph of $f(x)$ is plotted against x the derivative measures the slope of this graph at point x.

When $f(x)$ is a linear function of x, see Definition 15, an exact or constant value for the slope of the straight line is obtained.

If the function $f(x)$ is nonlinear, then the change in $f(x)$ divided by the change in x varies and the derivative determines an exact value for this rate of change at any given value of x. Therefore, this kind of functions is going to be analysed in more detail in this work.

A general concept, which provides a convenient method of stating background assumptions for future definitions, theorems, etc, is presented in the following definition.

Definition 18 (Factorial Function)

Let n be a positive integer. The factorial function of n is the product of all positive integers less than or equal to n. This function is denoted by $n!$ and defined by:

$$n! = \prod_{k=1}^{n} k \qquad \forall n \in \mathbb{N},$$

or recursively defined by:

$$n! = \begin{cases} 1 & \text{if } n = 0, \\ n(n-1)! & \text{if } n > 0 \end{cases} \qquad \forall n \in \mathbb{N}.$$

The double factorial *of a positive integer n is a generalization of the usual factorial $n!$. This is denoted by $n!!$ and defined recursively by:*

$$n!! = \begin{cases} 1 & \text{if } n = 0 \text{ or } n = 1, \\ n(n-2)!! & \text{if } n \geq 2 \end{cases}$$

The gamma function *is an extension of factorial to non-integer values of argument, e.g. w. For a complex number w with positive real part the* gamma function *is defined by:*

$$\Gamma(w) = \int_0^{\infty} t^{w-1} e^{-t} dt.$$

If n is a positive integer, then this function is defined by: $\Gamma(n) = (n-1)!$.

Assume a $n \times p$ data matrix, as given above, is given. The *arithmetical mean*, a reasonable estimator of the population mean $\boldsymbol{\mu}$, is presented in the following definition.

A.1.2 Basic Properties of Estimators

Let $x_1, x_2, \cdots, x_n$ be a realisation of an i.i.d. random sample drawn from a given probability distribution f_X with an unknown real parameter θ taking values in a *parameter space*, say $\mathfrak{D} \subset \mathbb{R}$.

A real-valued statistic (a function of the observable sample data) $T^*(x_1, x_2, \cdots, x_n)$ that is used to estimate an unknown population parameter θ is called, appropriately enough, an estimator of θ. Since an estimator is a random variable, it has a distribution and commonly a mean, a variance, and so on. These properties are introduced in the following definitions. For all these definitions the following assumption is made:

Let $x_1, x_2, \cdots, x_n$ be a realisation of an i.i.d. random sample drawn from a probability distribution with parameter θ. Furthermore, let T^* be an estimator for θ.

Definition 19 ((Random) Error)

The (random) error *is defined as the difference between the estimator and the true value of the parameter, i.e.*

$$\mathrm{RE}(T^*) = T^* - \theta.$$

The expected value of the (random) error is known as the *bias*.

Definition 20 (Bias of Estimators)

The bias *of* T^* *is defined as:* $\mathrm{b}(T^*, \theta) = \mathrm{E}[T^*] - \theta$, *i.e.*
the expected value of the estimator T^* *minus the true value of the parameter* θ.

This may be rewritten as: $\mathrm{b}(T^*, \theta) = \mathrm{E}[T^* - \theta]$, *i.e.*
the expected value of the difference between the estimator and the true value of the parameter, since the expected value of θ *is precisely* θ.

T^* *is an unbiased estimator of* θ *if the bias is zero.*

A measure of statistical dispersion of an estimator is obtained by averaging the squared distance of its possible values from the expected value (mean).

Definition 21 (Variance of Estimators)

The variance of T^* is defined as:

$$\text{var}(T^*) = \text{var}(T^*(x_1, x_2, \cdots, x_n)) = \text{E}[(T^* - \text{E}[T^*])^2].$$

The quality or performance of estimators is usually measured by computing the *Mean Squared Error*, i.e. in terms of its variation and unbiasedness.

Definition 22 (Mean Squared Error (MSE))

The Mean Squared Error *of T^* is defined as the expected value of the squared difference between the estimator and the true value of the parameter, i.e.*

$$\text{MSE}(T^*, \theta) = \text{E}[(T^* - \theta)^2].$$

The following property of the MSE holds true:

$$\text{MSE}(T^*, \theta) = \text{var}(T^*) + [\text{b}(T^*, \theta)]^2.$$

In particular, if the estimator is unbiased, then the MSE of T^* is simply the variance of T^*. In general, it is desired to have unbiased estimators with small MSE (small variance). Additionally, if two unbiased estimators of θ are obtained, denoted T_1^* and T_2^*, naturally the one with the smaller variance should be preferred.

A very useful asymptotic property of estimators is that of *consistency.* This property is introduced in the following definition.

Definition 23 (Consistency of Estimators)

Let $x_1, x_2, \cdots, x_n$ be a realisation of an i.i.d. random sample drawn from a probability distribution with parameter θ and T^* an estimator for θ. T^* is consistent as an estimator of θ if:

$$\lim_{n\to\infty} P[|T^*(x_1, x_2, \cdots, x_n) - \theta| \leq \epsilon) = 1 \ \forall \ \epsilon > 0.$$

It indicates that the estimator T^* will perform better and better as the sample size, say n, increases.

Remark A.1 (Consistency of Unbiased Estimators)

An unbiased estimator T^* is consistent if $\lim\limits_{n\to\infty} \text{var}(T^*(x_1, x_2, \cdots, x_n)) = 0.$

The following definitions are concerned with two well-known estimation methods in statistics.

Definition 24 (Maximum Likelihood (ML) Estimator)

Let $x_1, x_2, \cdots, x_n$ be a realisation of an i.i.d. random sample drawn from a probability distribution with parameter θ.

As a function of θ with $x_1, x_2, \cdots, x_n$ fixed, the likelihood function is given by:

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} f_X(x_i).$$

The method of maximum likelihood estimates θ by finding the value of θ that maximizes $\mathcal{L}(\theta)$. Thus, the maximum likelihood (ML) estimator of θ is given by:

$$\widehat{\theta} = \arg\max_{\theta} \mathcal{L}(\theta).$$

The maximum likelihood estimator is consistent. However, this estimator may not be unique, or indeed may not even exist. In general maximum likelihood estimators have desirable mathematical and optimality properties, for further details refer to Lehmann (1983).

In the following definition an unbiased estimator that has lower variance than any other unbiased estimator for all possible values of the parameter is introduced.

Definition 25 (Minimum-Variance Unbiased Estimator (MVUE))

Let $x_1, x_2, \cdots, x_n$ be a realisation of an i.i.d. random sample drawn from a probability distribution with parameter θ. Moreover, let $T^*(x_1, x_2, \cdots, x_n)$ be an estimation of θ, where $\theta \in \mathfrak{D}$ and $\mathfrak{D}$ is the parameter space. An unbiased estimator of θ is UMVU if $\forall$ $\theta \in \mathfrak{D}$ the following identity holds true:

$$\operatorname{var}(T^*(x_1, x_2, \cdots, x_n)) \leq \operatorname{var}(\tilde{T}^*(x_1, x_2, \cdots, x_n)),$$

for any other unbiased estimator $\tilde{T}^*$.

In Section A.1.3 an estimator approach which enables bias reduction has been explained in detail.

A.1.3 Generalised Jackknife Estimator

The generalized Jackknife estimator, or simply the Jackknife is an estimator, introduced by Quenouille (1956) for the purpose of bias reduction.

This estimation approach is presented as follows, for concepts and definitions refer to Section A.1.2.

Let $x_1, x_2, \cdots, x_n$ be a realisation of an i.i.d. random sample drawn from a probability distribution with parameter θ. Moreover, suppose that two functions, T_1^* and T_2^*, are defined over the n observations and are considered as two different estimators of the parameter θ.

Further suppose that each of these estimators is biased such that:

$$\mathrm{E}[T_i^*(x_1, x_2, \cdots, x_n)] - \theta = \mathrm{b}(T_i^*, \theta) \neq 0, i = 1, 2. \quad \text{(A.1.1)}$$

As mentioned in Schucany et al. (1971) the combination of the two estimators T_1^* and T_2^* may produce a third random variable which will often be an unbiased estimator for θ, conditions for unbiasedness will be presented in Remark A.2.

Let $\quad R = \dfrac{\mathrm{b}(T_1^*, \theta)}{\mathrm{b}(T_2^*, \theta)}$,

with $\mathrm{b}(T_1^*, \theta)$ and $\mathrm{b}(T_2^*, \theta)$ the respective bias terms of T_1^* and T_2^* with respect to θ.

The generalised Jackknife is given by:

$$T_3^* = G(T_1^*, T_2^*) = \frac{T_1^* - RT_2^*}{1 - R}. \quad \text{(A.1.2)}$$

The properties of the new generated estimator will be summarised in the following remark.

Remark A.2

If R depends on $1/n^m$ (for $m \geq 1$) and $\lim_{n\to\infty} R$ exists and is different from 1, then Schucany et al. (1971), Theorem 2.1, present two important properties of $G(T_1^*, T_2^*)$:

1. *when T_1^* and T_2^* are consistent for θ, then T_3^* is also consistent.*
2. *the quantity T_3^* is an unbiased estimator for θ.*

The variance of the new estimator depends jointly upon the value of R and the covariance between the two estimators T_1^* and T_2^*, as well as their variances. This is shown by:

$$\mathrm{var}(T_3^*) = \frac{1}{(1-R)^2}[\mathrm{var}(T_1^*) + R^2\mathrm{var}(T_2^*) - 2R\mathrm{cov}(T_1^*, T_2^*)]. \quad \text{(A.1.3)}$$

As pointed out in Schucany et al. (1971), "within the class of estimators for which R is fixed and positive we would desire that T_1^* and T_2^* have a high positive correlation. On the other hand it would appear that in the set of all $G(T_1^*, T_2^*)$ one should prefer to have $R < 0$ and T_1^* and T_2^* negatively correlated. Unfortunately a general method for accomplishing the latter is yet to be established".

A.1.4 Moments of Different Probability Distribution Functions

As pointed out in Mood, Graybill and Boes (1974), an additional way of characterising the *position* and *shape* of a probability distribution is by means of its *moments*. Moments are *expectations* of particular functions in the variable X. Since higher order moments will be used in this work, the *r-th central moments* as well as the *r-th non-central moments* are presented in the following definitions.

Definition 26 (Non-central Moments)

Let X be a random variable. The r-th non-central moment *of X, denoted by μ'_r, is defined by:*

(i) $\mathrm{E}[X^r] = \mu'_r = \sum_{j=1}^{N} x_j^r f_X(x_j)$, $j = 1, \cdots, N$, *if X is discrete.*

(ii) $\mathrm{E}[X^r] = \mu'_r = \int_{-\infty}^{\infty} x^r \, f_X(x)\, dx$, *if X is continuous.*

Notice that $\mu'_1 = \mathrm{E}[X]$, i.e. the mean of X.

Definition 27 (Central Moments)

Let X be a random variable. The r-th moment *of X about $\mathfrak{a}$ is defined by* $\mathrm{E}[(X-\mathfrak{a})^r]$. *The* rth moment *of X about $\mathfrak{a} = \mu$ is called the* rth central moment *of X, it is defined as follows:*

(i) $E[(X-\mu)^r] = \mu_r = \sum_{j=1}^{N} (x_j - \mu)^r f_X(x_j)$, $j = 1, \cdots, N$, *if X is discrete.*

(ii) $E[(X-\mu)^r] = \mu_r = \int_{-\infty}^{\infty} (x-\mu)^r\, f_X(x)\, dx$, *if* X *is continuous.*

Notice that the first four central moments have immediate interpretations:

- The first central moment is zero, i.e. $\mu_1 = \mathrm{E}[(X-\mu)] = 0$.
- The second central moment equals the variance of X, i.e. $\mu_2 = \mathrm{E}[(X-\mu)^2] = \mathrm{var}(X)$.
- The third and fourth moments about the mean are used to define the *standardized moments* which are used to define *skewness* and *kurtosis*, respectively.

The last two concepts are introduced in the following definitions.

Definition 28 (Skewness)

Let μ_3 *be the third central moment of a random variable* X *and* σ *its standard deviation. The skewness or third standardized moment is a measure of the asymmetry of the probability distribution of a random variable which is defined as:*

$$Sk = \frac{\mu_3}{\sigma^3}.$$

In probability theory and statistics, a measure of the "peakedness" of the probability distribution of a random variable is known as *kurtosis*.

This measure is defined more formally as follows.

Definition 29 (Kurtosis)

Let μ_4 *be the fourth central moment of a random variable* X *and* σ *its standard deviation.*
Kurtosis is defined as the fourth central moment divided by the standard deviation the power of 4 of the probability distribution minus 3, i.e.:

$$Ku = \frac{\mu_4}{\sigma^4} - 3.$$

Higher kurtosis means more of the variance is due to infrequent extreme deviations.

Remark A.3 (Relation Between Non-central and Central Moments)

Non-central moments can be converted to central moments. The general equation for converting the r-th order moment about the origin to the moment about the mean is given as:

$$\mu_r = \sum_{j=0}^{r} \binom{r}{j} (-1)^{r-j} \mu'_j \mu^{r-j},$$

where μ is the mean of the distribution and μ'_j the j-th moment about the origin.

For the particular cases $r = 2, 3, 4,$ which are of most interest because of the aforementioned relations to variance, skewness and kurtosis, respectively, the expression μ_r becomes:

$$\mu_2 = \mu'_2 - \mu^2,$$
$$\mu_3 = \mu'_3 - 3\mu\mu'_2 + 2\mu^3,$$
$$\mu_4 = \mu'_4 - 4\mu\mu'_3 + 6\mu^2\mu'_2 - 3\mu^4.$$

In Section 4.1 it was pointed out that for the calculation of $\text{var}(T_1^{Poly})$ and $\text{var}(T_2^{Poly})$ the third and fourth non-central and central moments are needed. In this section, the corresponding moments of different established continuous probability distributions, such as the normal distribution, uniform distribution, exponential distribution, lognormal distribution and gamma distribution, as well as an established discrete probability distribution, such as the Bernoulli distribution, will be presented in more detail. For definitions of random variables and of probability distribution functions refer Mood, Graybill and Boes (1974).

Expressions for the first four non-central as well as central moments of a normally distributed variable X are listed in the following table:

Distribution	Non-Central Moments	Central Moments
Normal	$\mu'_r = \frac{1}{\sigma\sqrt{(2\pi)}} \int_{-\infty}^{\infty} x^r \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx$	
	$\mu'_1 = \mu$	$\mu_1 = 0$
	$\mu'_2 = \mu^2 + \sigma^2$	$\mu_2 = \sigma^2$
	$\mu'_3 = \mu(\mu^2 + 3\sigma^2)$	$\mu_3 = 0$
	$\mu'_4 = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$	$\mu_4 = 3\sigma^4$
Exponential	$\mu'_r = \lambda^{-r} r!$	
	$\mu'_1 = 1/\lambda$	$\mu_1 = 0$
	$\mu'_2 = 2/\lambda^2$	$\mu_2 = 1/\lambda^2$
	$\mu'_3 = 6/\lambda^3$	$\mu_3 = 2/\lambda^3$
	$\mu'_4 = 24/\lambda^4$	$\mu_4 = 9/\lambda^4$
Uniform	$\mu'_r = \frac{b^{r+1} - a^{r+1}}{(r+1)(b-a)}$	
	$\mu'_1 = (1/2)(a+b)$	$\mu_1 = 0$
	$\mu'_2 = (1/3)(a^2 + ab + b^2)$	$\mu_2 = (1/12)(b-a)^2$
	$\mu'_3 = (1/4)(a+b)(a^2 + b^2)$	$\mu_3 = 0$
	$\mu'_4 = (1/5)(a^4 + a^3 b + a^2 b^2 + ab^3 + b^4)$	$\mu_4 = (1/80)(b-a)^4$
Lognormal	$\mu'_r = \exp\left[r\mu + \frac{r^2\sigma^2}{2}\right]$	
	$\mu'_1 = \exp[\mu + \frac{\sigma^2}{2}]$	$\mu_1 = 0$
	$\mu'_2 = \exp[2(\mu + \sigma^2)]$	$\mu_2 = \exp[2\mu + \sigma^2](\exp[\sigma^2] - 1)$
	$\mu'_3 = \exp[3\mu + 9\frac{\sigma^2}{2}]$	$\mu_3 = \exp[3\mu + 3\frac{\sigma^2}{2}]$
		$(\exp[\sigma^2] - 1)^2(\exp[\sigma^2] + 2)$
	$\mu'_4 = \exp[4\mu + 8\sigma^2]$	$\mu_4 = \exp[4\mu + 2\sigma^2](\exp[\sigma^2] - 1)^2$
		$+ \ (\exp[4\sigma^2]2\exp[3\sigma^2]$
		$+ \ 3\exp[2\sigma^2] - 3)$
Gamma	$\mu'_r = \frac{\beta^r \Gamma(\alpha + r)}{\Gamma(\alpha)}$	
	$\mu'_1 = \beta\alpha$	$\mu_1 = 0$
	$\mu'_2 = \beta^2(\alpha+1)\alpha$	$\mu_2 = \alpha$
	$\mu'_3 = \beta^3(\alpha+2)(\alpha+1)\alpha$	$\mu_3 = 2\alpha$
	$\mu'_4 = \beta^4(\alpha+3)(\alpha+2)(\alpha+1)\alpha$	$\mu_4 = 3\alpha^2 + 6\alpha$
Bernoulli	$\mu'_r = p$	
	$\mu'_1 = p$	$\mu_1 = 0$
	$\mu'_2 = p$	$\mu_2 = p(1-p)$
	$\vdots$	$\mu_3 = p(1-p)(1-2p)$
	$\vdots$	$\mu_4 = p(1-p)(3p^2 - 3p + 1)$

Table A.1: Non-Central and Central Moments of Different Distributions

A.2 Additional Remarks

Remark A.4 (Mean and Variance of T_1^{Poly} and T_2^{Poly})

Suppose $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ *represents a realisation of an i.i.d. random sample drawn from a* p*–dimensional probability distribution with mean* $\boldsymbol{\mu} = \mathrm{E}[\boldsymbol{x}_i] = \boldsymbol{g}$ *and covariance matrix* $\boldsymbol{\Sigma} = \mathrm{E}[(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})']$*. Then, consider* $\boldsymbol{z}_i = \boldsymbol{x}_i - \boldsymbol{\mu}$, $i = 1, \ldots, n$, *with moments given as in Appendix A.5.2 and in Remark A.1. The elements of* $\boldsymbol{z}_i$ *are* z_{ij}*, for* $j = 1, \cdots, p$*, with* $\mathrm{E}[z_{ij}] = 0$, $\mathrm{E}[z_{ij}^2] = \sigma_j^2$, $\mathrm{E}[z_{ij}^3] = \eta_j \sigma_j^3$ *and* $\mathrm{E}[z_{ij}^4] = \upsilon_j \sigma_j^4$*.*

Furthermore, consider the nonstochastic vectors $\boldsymbol{a}$ *and* $\boldsymbol{b}$ *and the nonstochastic symmetric matrices* $\boldsymbol{A}$ *and* $\boldsymbol{B}$*. The following identities hold true:*

(i) $\Psi(\boldsymbol{A}) = (tr\boldsymbol{A\Sigma})\boldsymbol{\Sigma} + 2\boldsymbol{\Sigma A \Sigma} + \boldsymbol{\Sigma}\mathrm{Diag}(\boldsymbol{A})\boldsymbol{\Delta\Sigma}$

(ii) $\Phi^*(\boldsymbol{A}) = \boldsymbol{\Sigma}^{3/2}\mathrm{Diag}(\boldsymbol{A})\boldsymbol{\eta}$

(iii) $\Phi(\boldsymbol{a}) = \mathrm{Diag}(\boldsymbol{a}\boldsymbol{\eta}')\boldsymbol{\Sigma}^{3/2}$

(iv) $tr\boldsymbol{A}\Psi(\boldsymbol{A}) = \alpha + 2\beta + \boldsymbol{A\Sigma}\mathrm{Diag}(\boldsymbol{A})\boldsymbol{\Delta\Sigma}$

(v) $tr\boldsymbol{A}\Psi(\boldsymbol{B}) = tr\boldsymbol{B}\Psi(\boldsymbol{A}) = tr(\boldsymbol{A} \otimes \boldsymbol{B})\boldsymbol{\Psi}$

(vi) $\boldsymbol{a}'\Phi^*(\boldsymbol{A}) = tr\boldsymbol{A}\Phi(\boldsymbol{a}) = tr(\boldsymbol{a}' \otimes \boldsymbol{A})\boldsymbol{\Phi}$

(vii) $\mathrm{E}[\boldsymbol{a}'\boldsymbol{x}_i + \boldsymbol{x}_i'\boldsymbol{A}\boldsymbol{x}_i] = \boldsymbol{a}'\boldsymbol{g} + \boldsymbol{g}'\boldsymbol{A}\boldsymbol{g} + tr\boldsymbol{AV}$

(viii) $$\begin{aligned}\mathrm{var}(\boldsymbol{a}'\boldsymbol{x}_i + \boldsymbol{x}_i'\boldsymbol{A}\boldsymbol{x}_i) \;=\; & 4\boldsymbol{g}'\boldsymbol{AVAg} + 4tr(\boldsymbol{g}'\boldsymbol{A} \otimes \boldsymbol{A})\boldsymbol{\Phi} + tr(\boldsymbol{A} \otimes \boldsymbol{A})\boldsymbol{\Psi} \\ & + 4\boldsymbol{a}'\boldsymbol{VAg} + 2tr(\boldsymbol{a}' \otimes \boldsymbol{A})\boldsymbol{\Phi} + \boldsymbol{a}'\boldsymbol{Va} - (tr\boldsymbol{AV})^2.\end{aligned}$$

(ix) $$\begin{aligned}\mathrm{cov}(\boldsymbol{a}'\boldsymbol{x}_i + \boldsymbol{x}_i'\boldsymbol{A}\boldsymbol{x}_i, \boldsymbol{b}'\boldsymbol{x}_i + \boldsymbol{x}_i'\boldsymbol{B}\boldsymbol{x}_i) \;=\; & \boldsymbol{b}'[2\boldsymbol{\Sigma A g} + \boldsymbol{\Sigma a} + \Phi^*(\boldsymbol{A})] \\ & + tr\boldsymbol{B}[4\boldsymbol{g}\boldsymbol{g}'\boldsymbol{A\Sigma} + 2\Phi(\boldsymbol{Ag}) + 2\Phi^*(\boldsymbol{A})\boldsymbol{g}' \\ & + \Psi(\boldsymbol{A}) + 2\boldsymbol{g}\boldsymbol{a}'\boldsymbol{\Sigma} + \Phi(\boldsymbol{a}) - (tr\boldsymbol{A\Sigma})\boldsymbol{\Sigma}],\end{aligned}$$

where $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_p)'$, $\mathrm{Diag}(\boldsymbol{A})$ *is a diagonal matrix obtained from* $\boldsymbol{A}$ *by replacing the off-diagonal elements of* $\boldsymbol{A}$ *by zero,* $\boldsymbol{\Delta}$ *is a diagonal matrix with* $\upsilon_1 - 3, \cdots, \upsilon_p - 3$ *as the diagonal elements and tr the trace of a square matrix. Notice that* Ψ *and* Φ *are linear functions instead of matrices,* Ψ *from* $\mathbb{R}^p \to \mathcal{M}_{p\times p}$ *and* Φ *from* $\mathcal{M}_{p\times p} \to \mathcal{M}_{p\times p}$*, where* $\mathcal{M}_{p\times p}$ *stands for the set of all symmetric* $p \times p$ *matrices,* Φ^* *is the conjugated or transposed operator with respect to the usual inner product of matrices.*

Remark A.5 (New Approach Based on Kleffe and Rao (1988, Section 2.1))

Kleffe and Rao (1988, Theorem 2.1.2) have shown equivalent expressions to the equations (i) - (iii)*, from Remark A.4, for* **correlated** *variables* X_1 *and* X_2*, i.e.* $p = 2$*. These expressions are given as follows:*

(i) $\Psi_{\boldsymbol{z}}(\boldsymbol{A}) = (tr\boldsymbol{A\Sigma})\boldsymbol{\Sigma} + 2\boldsymbol{\Sigma A \Sigma} + \mathrm{Diag}(\boldsymbol{D}_1, \ldots, \boldsymbol{D}_n)$,

with $\boldsymbol{D}_i = \Psi_i(\boldsymbol{A}_{ii}) - 2\boldsymbol{\Sigma A}_{ii}\boldsymbol{\Sigma} - (tr\boldsymbol{A}_{ii}\boldsymbol{\Sigma})\boldsymbol{\Sigma}$, $i = 1, \cdots, n$*. Since* $\boldsymbol{A}_{ii} = \boldsymbol{A}$*, then* $\boldsymbol{D}_i = \Psi_u(\boldsymbol{A}) - 2\boldsymbol{\Sigma A \Sigma} - (tr\boldsymbol{A\Sigma})\boldsymbol{\Sigma}$ *and* $\Psi_u(\boldsymbol{A}) = \sum\limits_{i=1}^{p}\sum\limits_{j=1}^{p} A_{ij}\boldsymbol{\Psi}_{ij}$.

(ii) $\Phi^*_{\boldsymbol{z}}(\boldsymbol{A}) = (\Phi^*_1(\boldsymbol{A}_{11})', \ldots, \Phi^*_n(\boldsymbol{A}_{nn})')'$, *with* $\Phi^*_i(\boldsymbol{A}_{ii}) = \Phi^*_u(\boldsymbol{A}) = (tr\boldsymbol{A\Phi}_1, \ldots, tr\boldsymbol{A\Phi}_p)$.

(iii) $\Phi_{\boldsymbol{z}}(\boldsymbol{a}) = \mathrm{Diag}(\Phi_1(\boldsymbol{a}_1), \ldots, \Phi_n(\boldsymbol{a}_n))$, *with* $\Phi_i(\boldsymbol{a}_i) = \Phi_u(\boldsymbol{a}) = \sum\limits_{i=1}^{p} a_i\boldsymbol{\Phi}_i$.

Remark A.6 (Central Moments of two Correlated Variables)

Let $\boldsymbol{\mathcal{Y}} = (Y_1, Y_2)'$ *and* $\boldsymbol{\mathcal{Z}} = (Z_1, Z_2)'$ *be given as in Section 6.1.1. The first four central moments of the correlated random variables* Y_j*, i.e. non-central moments of* Z_j *are:*

First central moments of Y_j*, for* $j = 1, 2$*:*

$$\mathrm{E}[Z_j] = \mathrm{E}[Y_j - \mu_1] = \mathrm{E}[Y_j] - \mu_j = 0.$$

Second central moments of Y_j*, for* $j = 1, 2$*:*

$$\mathrm{E}[Z_j^2] = \mathrm{E}[Y_j - \mu_j]^2 = \mathrm{var}(Y_j) = \mathrm{E}[\tau_j X_1 + \tau_j q X_2]^2 = \tau_j^2\mathrm{E}[X_1^2] + \tau_j^2 q^2 \mathrm{E}[X_2^2].$$

Since $\mathrm{E}[X_1^2] = \mathrm{E}[X_2^2] = 1$ *it follows:* $\mathrm{E}[Z_j^2] = \tau_j^2(1+q^2) = \frac{\sigma_j^2}{(1+q^2)}(1+q^2) = \sigma_j^2$.

Third central moments of Y_j*, for* $j = 1, 2$*:*

$$\Phi_j = \mathrm{E}[Z_j^3] = \mathrm{E}[Y_j - \mu_j]^3 = \mathrm{E}[\tau_j X_1 + \tau_j q X_2]^3 = \tau_j^3\mathrm{E}[X_1^3] + \tau_j^3 q^3 \mathrm{E}[X_2^3].$$

Since $\mathrm{E}[X_1^3] = \mathrm{E}[X_2^3] = \mathrm{E}[X_j^3]$ *it follows:* $\Phi_j = \mathrm{E}[X_j^3]\tau_j^3(1+q^3)$.

Fourth central moments of Y_j*, for* $j = 1, 2$*:*

$$\Psi_j = \mathrm{E}[Z_j^4] = \mathrm{E}[Y_j - \mu_j]^4 = \mathrm{E}[\tau_j X_1 + \tau_j q X_2]^4 = \tau_j^4\mathrm{E}[X_1^4] + 6\mathrm{E}[\tau_j^2 X_1^2 \tau_j^2 q^2 X_2^2] + \tau_j^4 q^4 \mathrm{E}[X_2^4].$$

Since $\mathrm{E}[X_1^4] = \mathrm{E}[X_2^4] = \mathrm{E}[X_j^4]$ *and* $\mathrm{E}[X_1^2] = \mathrm{E}[X_2^2] = 1$ *it follows:*

$$\Psi_j = \mathrm{E}[Z_j^4] = \tau_j^4[\mathrm{E}[X_j^4](1+q^4) + 6q^2], \text{ with } \tau_j = \sqrt{\frac{\sigma_j^2}{1+q^2}},\ j = 1, 2 \text{ and } q = \frac{1-\sqrt{1-\rho^2}}{\rho},$$

for $\rho \neq 0$ *and* $q = 0$*, for* $\rho = 0$.

It can be seen that for $q = 0$ these expressions are equivalent to those given for uncorrelated random variables, see Appendix A.5.2.

A.3 Additional Lemmas

Lemma A.1 (Moment Vectors and Matrices)

Suppose $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ *represents a realisation of an i.i.d. random sample drawn from a p–dimensional probability distribution with mean* $\boldsymbol{\mu} = E[\boldsymbol{x}_i]$ *and covariance matrix* $\boldsymbol{\Sigma} = \mathrm{E}[(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})']$. *Furthermore, consider the random vectors* $\boldsymbol{z}_i = \boldsymbol{x}_i - \boldsymbol{\mu}$, $i = 1, \ldots, n$, *with existing moments* $\mathrm{E}[\boldsymbol{z}_i] = \mathbf{0}$, $\mathrm{E}[\boldsymbol{z}_i \boldsymbol{z}_i'] = \boldsymbol{\Sigma}$, $\mathrm{E}[\boldsymbol{z}_i \otimes \boldsymbol{z}_i \boldsymbol{z}_i'] = \boldsymbol{\Phi}$ *and* $\mathrm{E}[\boldsymbol{z}_i \boldsymbol{z}_i' \otimes \boldsymbol{z}_i \boldsymbol{z}_i'] = \boldsymbol{\Psi}$, *with* $\boldsymbol{z}_i$, $i = 1 \cdots, n$ **independent** *vectors.*

Now, let $\boldsymbol{y} = \mathrm{Vec}(\boldsymbol{X}')$ *be given as in Remark 3.2 and* $\boldsymbol{z} = \boldsymbol{y} - \mathrm{E}[\boldsymbol{y}]$, *with* $\mathrm{E}(\boldsymbol{y}) = \boldsymbol{g}_* = \mathbf{1}_n \otimes \boldsymbol{\mu}$. *Then the first moment vector and the second, third and fourth moment matrices of* $\boldsymbol{z}$ *are given by:*

$$\text{(i)} \quad \mathrm{E}[\boldsymbol{z}] = \mathbf{1}_n \otimes \mathbf{0}, \ \textit{with } \mathbf{0} \in \mathbb{R}^p$$

$$\text{(ii)} \quad \boldsymbol{V}_* = \mathrm{E}[\boldsymbol{z}\boldsymbol{z}'] = \mathrm{cov}(\boldsymbol{y}) = \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}$$

$$\text{(iii)} \quad \boldsymbol{\Phi}_* = \mathrm{E}[\boldsymbol{z} \otimes \boldsymbol{z}\boldsymbol{z}'] = (\boldsymbol{I}_n \otimes \boldsymbol{K}_{pn} \otimes \boldsymbol{I}_p)(\boldsymbol{G} \otimes \boldsymbol{\Phi}),$$

where $\boldsymbol{G} = (\boldsymbol{E}_{11}, \ldots, \boldsymbol{E}_{nn})'$, $\boldsymbol{E}_{ii} = \boldsymbol{e}_i \boldsymbol{e}_i'$, *with* $\boldsymbol{e}_i$ *being the* i*-th member of the canonical basis in* $\mathbb{R}^n$, *and* $\boldsymbol{K}_{pn} = \boldsymbol{K}_{p,n}$ *is the commutation matrix of type* $pn \times pn$.

$$\begin{aligned} \text{(iv)} \quad \boldsymbol{\Psi}_* = \ & \mathrm{E}[\boldsymbol{z}\boldsymbol{z}' \otimes \boldsymbol{z}\boldsymbol{z}'] = (\boldsymbol{I}_{n^2p^2} + \boldsymbol{K}_{np,np})(\boldsymbol{I}_n \otimes \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}) \\ & + [\mathrm{Vec}\ (\boldsymbol{I}_n \otimes \boldsymbol{\Sigma})][\mathrm{Vec}\ (\boldsymbol{I}_n \otimes \boldsymbol{\Sigma})]' + (\boldsymbol{I}_n \otimes \boldsymbol{K}_{pn} \otimes \boldsymbol{I}_p) \cdot \\ & \{\widetilde{\boldsymbol{K}}_{nn} \otimes [\boldsymbol{\Psi} - (\mathrm{Vec}\ \boldsymbol{\Sigma})(\mathrm{Vec}\ \boldsymbol{\Sigma})' - (\boldsymbol{I}_{p^2} + \boldsymbol{K}_{pp})(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})]\} \cdot \\ & (\boldsymbol{I}_n \otimes \boldsymbol{K}_{np} \otimes \boldsymbol{I}_p), \end{aligned}$$

where $\boldsymbol{K}_{np,np}$ *is the commutation matrix of type* $n^2p^2 \times n^2p^2$ *and* $\widetilde{\boldsymbol{K}}_{nn} = \sum_{i=1}^{n} (\boldsymbol{E}_{ii} \otimes \boldsymbol{E}_{ii})$.

For proof see Neudecker and Trenkler (2002) and for properties of the commutation matrix see Appendix A.4.3.

Lemma A.2 (Mean and Variance of T_1^{Poly} and T_2^{Poly} for whole Multivariate Sample)

Consider $\boldsymbol{z} = \boldsymbol{y} - \mathrm{E}[\boldsymbol{y}]$, *with* $\boldsymbol{y} = \mathrm{Vec}\boldsymbol{X}'$ *and* $\mathrm{E}(\boldsymbol{y}) = \boldsymbol{g}_* = \mathbf{1}_n \otimes \boldsymbol{\mu}$. *Then* $\mathrm{E}[\boldsymbol{z}] = \mathbf{1}_n \otimes \mathbf{0}$ *and* $\boldsymbol{V}_* = \mathrm{E}[\boldsymbol{z}\boldsymbol{z}'] = \mathrm{cov}(\boldsymbol{y}) = \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}$.

The following identities can be deduced for the approximations T_1^{Poly} *and* T_2^{Poly} *by using Remark A.4:*

(i.1) $\Psi(\boldsymbol{F}_1) = \frac{1}{n}[tr\boldsymbol{A}\boldsymbol{\Sigma}(\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}) + \frac{2}{n}\mathbf{1}_n\mathbf{1}_n' \otimes \boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\Sigma} + \frac{1}{n}\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}\mathrm{Diag}(\boldsymbol{A})\boldsymbol{\Delta}\boldsymbol{\Sigma}]$

(i.2) $\Psi(\boldsymbol{F}_2) = [tr\boldsymbol{A}\boldsymbol{\Sigma}(\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}) + \frac{2}{n}\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\Sigma} + \frac{1}{n}\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}\mathrm{Diag}(\boldsymbol{A})\boldsymbol{\Delta}\boldsymbol{\Sigma}]$

(ii.1) $\Phi^*(\boldsymbol{F}_1) = \frac{1}{n^2}\mathbf{1}_n \otimes \boldsymbol{\Sigma}^{3/2}\mathrm{Diag}(\boldsymbol{A})\boldsymbol{\eta}$

(ii.2) $\Phi^*(\boldsymbol{F}_2) = \frac{1}{n}\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}^{3/2}\mathrm{Diag}(\boldsymbol{A})\boldsymbol{\eta}$

(iii) $\Phi(\boldsymbol{f}) = \frac{1}{n}\boldsymbol{I}_n \otimes \mathrm{Diag}(\boldsymbol{a}\boldsymbol{\eta}')\boldsymbol{\Sigma}^{3/2}$

(iv.1) $tr\boldsymbol{F}_1\Psi(\boldsymbol{F}_1) = tr(\boldsymbol{F}_1 \otimes \boldsymbol{F}_1)\boldsymbol{\Psi}_* = \frac{1}{n^3}[tr\boldsymbol{A}\Psi(\boldsymbol{A}) + (n-1)(2\beta + \alpha)]$

(iv.2) $tr\boldsymbol{F}_2\Psi(\boldsymbol{F}_2) = tr(\boldsymbol{F}_2 \otimes \boldsymbol{F}_2)\boldsymbol{\Psi}_* = \frac{1}{n}[tr\boldsymbol{A}\Psi(\boldsymbol{A}) + (n-1)\alpha]$

(v) $tr\boldsymbol{F}_2\Psi(\boldsymbol{F}_1) = tr(\boldsymbol{F}_2 \otimes \boldsymbol{F}_1)\boldsymbol{\Psi}_* = \frac{1}{n^2}[tr\boldsymbol{A}\Psi(\boldsymbol{A}) + (n-1)\alpha]$

(vi.1) $\boldsymbol{f}'\Phi^*(\boldsymbol{F}_1) = tr(\boldsymbol{F}_1)\Phi(\boldsymbol{f}) = tr(\boldsymbol{f}' \otimes \boldsymbol{F}_1)\boldsymbol{\Phi}_* = \frac{1}{n^2}tr(\boldsymbol{a}' \otimes \boldsymbol{A})\boldsymbol{\Phi} = \frac{1}{n^2}tr\boldsymbol{A}\Phi(\boldsymbol{a})$

(vi.2) $\boldsymbol{f}'\Phi^*(\boldsymbol{F}_2) = tr(\boldsymbol{F}_2)\Phi(\boldsymbol{f}) = tr(\boldsymbol{f}' \otimes \boldsymbol{F}_2)\boldsymbol{\Phi}_* = \frac{1}{n}tr(\boldsymbol{a}' \otimes \boldsymbol{A})\boldsymbol{\Phi} = \frac{1}{n}tr\boldsymbol{A}\Phi(\boldsymbol{a})$

(vii) $\mathrm{E}[\boldsymbol{f}'\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{F}_l\boldsymbol{y}] = \boldsymbol{f}'\boldsymbol{g}_* + \boldsymbol{g}_*'\boldsymbol{F}_l\boldsymbol{g}_* + tr\boldsymbol{F}_l\boldsymbol{V}_*, l = 1, 2$

(viii)
$$\begin{aligned}\mathrm{var}(\boldsymbol{f}'\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{F}_l\boldsymbol{y}) \quad = \quad & 4\boldsymbol{g}_*'\boldsymbol{F}_l\boldsymbol{V}_*\boldsymbol{F}_l\boldsymbol{g}_* + 4tr(\boldsymbol{g}_*'\boldsymbol{F}_l \otimes \boldsymbol{F}_l)\boldsymbol{\Phi}_* \\ & + tr(\boldsymbol{F}_l \otimes \boldsymbol{F}_l)\boldsymbol{\Psi}_* + 4\boldsymbol{f}'\boldsymbol{V}_*\boldsymbol{F}_l\boldsymbol{g}_* \\ & + 2tr(\boldsymbol{f}' \otimes \boldsymbol{F}_l)\boldsymbol{\Phi}_* + \boldsymbol{f}'\boldsymbol{V}_*\boldsymbol{f} - (tr\boldsymbol{F}_l\boldsymbol{V}_*)^2\end{aligned}$$

(ix)
$$\begin{aligned}\mathrm{cov}(\boldsymbol{f}'\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{F}_1\boldsymbol{y}, \boldsymbol{f}'\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{F}_2\boldsymbol{y}) \quad = \quad & \boldsymbol{f}'[2\boldsymbol{V}_*\boldsymbol{F}_1\boldsymbol{g}_* + \boldsymbol{V}_*\boldsymbol{f} + \Phi^*(\boldsymbol{F}_1)] \\ & + tr\boldsymbol{F}_2[4\boldsymbol{g}_*\boldsymbol{g}_*'\boldsymbol{F}_1\boldsymbol{V}_* + 2\Phi(\boldsymbol{F}_1\boldsymbol{g}_*) \\ & + 2\Phi^*(\boldsymbol{F}_1)\boldsymbol{g}_*' + \Psi(\boldsymbol{F}_1) \\ & + 2\boldsymbol{g}_*\boldsymbol{f}'\boldsymbol{V}_* + \Phi(\boldsymbol{f}) - (tr\boldsymbol{F}_1\boldsymbol{V}_*)\boldsymbol{V}_*].\end{aligned}$$

Proof: See Appendix C.1.

Lemma A.3 (Approach by Kleffe and Rao (1988) for whole Multivariate Sample)

Let $\Psi_{\boldsymbol{z}}$ and $\Phi_{\boldsymbol{z}}$ be linear functions and $\Phi^*_{\boldsymbol{z}}$ the conjugated or transposed operator with respect to the usual inner product of matrices given as in Remark A.5. For **correlated** *random variables X_1 and X_2 the following identities, useful for the calculation of the variances and covariances, were obtained:*

(i.1) $\Psi_{\boldsymbol{z}}(\boldsymbol{F}_1) = \frac{1}{n}[(tr\boldsymbol{A\Sigma})(\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}) + \frac{2}{n}\boldsymbol{1}_n\boldsymbol{1}_n^{'} \otimes \boldsymbol{\Sigma A \Sigma} + \frac{1}{n}\boldsymbol{I}_n \otimes \mathrm{Diag}(\boldsymbol{D}_i)]$

(i.2) $\Psi_{\boldsymbol{z}}(\boldsymbol{F}_2) = (tr\boldsymbol{A\Sigma})(\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}) + \frac{2}{n}\boldsymbol{I}_n\boldsymbol{1}_n^{'} \otimes \boldsymbol{\Sigma A \Sigma} + \frac{1}{n}\boldsymbol{I}_n \otimes \mathrm{Diag}(\boldsymbol{D}_i)$

(ii.1) $\Phi^*_{\boldsymbol{z}}(\boldsymbol{F}_1) = \frac{1}{n^2}\boldsymbol{1}_n \otimes \Phi^*_u(\boldsymbol{A})$

(ii.2) $\Phi^*_{\boldsymbol{z}}(\boldsymbol{F}_2) = \frac{1}{n}\boldsymbol{1}_n \otimes \Phi^*_u(\boldsymbol{A})$

(iii) $\Phi_{\boldsymbol{z}}(\boldsymbol{f}) = \frac{1}{n}\boldsymbol{I}_n \otimes \mathrm{Diag}(\Phi_u(\boldsymbol{a}))$.

Proof: See Appendix C.3.

Lemma A.4 (Mean and Variance of T_1^{Poly} and T_2^{Poly} for whole Univariate Sample)

Assume that $\boldsymbol{z} = \boldsymbol{x} - \mathrm{E}[\boldsymbol{x}]$, with $\mathrm{E}[\boldsymbol{z}] = \boldsymbol{0}$, $\boldsymbol{g}_* = \mathrm{E}[\boldsymbol{x}] = \mu\boldsymbol{1}_n$ and $\boldsymbol{V}_* = \mathrm{E}[\boldsymbol{zz}'] = \mathrm{cov}(\boldsymbol{x}) = \sigma^2\boldsymbol{I}_n$. Furthermore, assume the elements of $\boldsymbol{z}$ are z_i, with $E[z_i] = 0$, $E[z_i^2] = \sigma^2$, $E[z_i^3] = \eta\sigma^3$ and $E[z_i^4] = \upsilon\sigma^4$. Using the expressions given in Remark A.4 and different matrix operations the following properties hold true:

(i.1) $\Psi(\boldsymbol{F}_1) = \frac{1}{n}[A\sigma^4\boldsymbol{I}_n + \frac{1}{n}2A\sigma^4\boldsymbol{1}_n\boldsymbol{1}_n^{'} + \frac{A\Delta\sigma^4}{n}\boldsymbol{I}_n]$

(i.2) $\Psi(\boldsymbol{F}_2) = [A\sigma^4\boldsymbol{I}_n + \frac{1}{n}2A\sigma^4\boldsymbol{I}_n + +\frac{A\Delta\sigma^4}{n}\boldsymbol{I}_n]$

(ii.1) $\Phi^*(\boldsymbol{F}_1) = \frac{1}{n^2}A\eta\sigma^3\boldsymbol{1}_n$

(ii.2) $\Phi^*(\boldsymbol{F}_2) = \frac{1}{n}A\eta\sigma^3\boldsymbol{I}_n$

(iii)$\Phi(\boldsymbol{f}) = \frac{1}{n}a\eta\sigma^3\boldsymbol{I}_n$

(iv.1)$tr\boldsymbol{F}_1\Psi(\boldsymbol{F}_1) = \frac{1}{n^3}[tr\boldsymbol{A}\Psi(\boldsymbol{A}) + (n-1)(3\alpha)]$

(iv.2)$tr\boldsymbol{F}_2\Psi(\boldsymbol{F}_2) = \frac{1}{n}[tr\boldsymbol{A}\Psi(\boldsymbol{A}) + (n-1)\alpha]$

(v)$tr\boldsymbol{F}_2\Psi(\boldsymbol{F}_1) = \frac{1}{n^2}[tr\boldsymbol{A}\Psi(\boldsymbol{A}) + (n-1)\alpha]$

(vi.1)$\boldsymbol{f}'\Phi^*(\boldsymbol{F}_1) = tr(\boldsymbol{F}_1)\Phi(\boldsymbol{f}) = \frac{1}{n^2}a\eta\sigma^3 A$

(vi.2)$\boldsymbol{f}'\Phi^*(\boldsymbol{F}_2) = tr(\boldsymbol{F}_2)\Phi(\boldsymbol{f}) = \frac{1}{n}a\eta\sigma^3 A$,

where $\boldsymbol{f} = \frac{a}{n}\boldsymbol{1}_n$, $\boldsymbol{F}_1 = \frac{A}{n^2}\boldsymbol{1}_n\boldsymbol{1}_n^{'}$, $\boldsymbol{F}_2 = \frac{A}{n}\boldsymbol{I}_n$ *and* $tr\boldsymbol{A}\Psi(\boldsymbol{A}) = 3\alpha + \boldsymbol{A\Sigma}\mathrm{Diag}\ (\boldsymbol{A})\boldsymbol{\Delta\Sigma}$.

Proof: See Appendix C.5

A.4 Matrix Operations

A.4.1 The Vec Operator

The Vec operator of an $m \times n$ matrix $\boldsymbol{A}$, denoted by Vec($\boldsymbol{A}$), is the $mn \times 1$ column vector obtained by stacking the column vectors of $\boldsymbol{A} = [\boldsymbol{a}_1 \boldsymbol{a}_2 \cdots \boldsymbol{a}_n]$ below one another. For example, for the 2×2 matrix $\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, the Vec operator is $\mathrm{Vec}(\boldsymbol{A}) = \begin{bmatrix} a_{11} \\ a_{12} \\ a_{21} \\ a_{22} \end{bmatrix}$.

For more details reference is made to Schmidt and Trenkler (2006).

A.4.2 The Kronecker Product

The Kronecker product, denoted by $\otimes$, is an operation on two matrices of arbitrary dimension. The result of this product is a block matrix.

If $\boldsymbol{A}$ is an $m \times n$ matrix and $\boldsymbol{B}$ is a $p \times q$ matrix, then the Kronecker product $\boldsymbol{A} \otimes \boldsymbol{B}$ is the $mp \times nq$ block matrix. It is represented as follows:

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{11}\boldsymbol{B} & \cdots & a_{1n}\boldsymbol{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\boldsymbol{B} & \cdots & a_{mn}\boldsymbol{B} \end{bmatrix}.$$

The product can be represented in further details, as follows:

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{11}b_{11} & a_{11}b_{12} & \cdots & a_{11}b_{1q} & \cdots & \cdots & a_{1n}b_{11} & a_{1n}b_{12} & \cdots & a_{1n}b_{1q} \\ a_{11}b_{21} & a_{11}b_{22} & \cdots & a_{11}b_{2q} & \cdots & \cdots & a_{1n}b_{21} & a_{1n}b_{22} & \cdots & a_{1n}b_{2q} \\ \vdots & \vdots & \ddots & \vdots & & & \vdots & \vdots & \ddots & \vdots \\ a_{11}b_{p1} & a_{11}b_{p2} & \cdots & a_{11}b_{pq} & \cdots & \cdots & a_{1n}b_{p1} & a_{1n}b_{p2} & \cdots & a_{1n}b_{pq} \\ \vdots & \vdots & & \vdots & \ddots & & \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & & \ddots & \vdots & \vdots & & \vdots \\ a_{m1}b_{11} & a_{m1}b_{12} & \cdots & a_{m1}b_{1q} & \cdots & \cdots & a_{mn}b_{11} & a_{mn}b_{12} & \cdots & a_{mn}b_{1q} \\ a_{m1}b_{21} & a_{m1}b_{22} & \cdots & a_{m1}b_{2q} & \cdots & \cdots & a_{mn}b_{21} & a_{mn}b_{22} & \cdots & a_{mn}b_{2q} \\ \vdots & \vdots & \ddots & \vdots & & & \vdots & \vdots & \ddots & \vdots \\ a_{m1}b_{p1} & a_{m1}b_{p2} & \cdots & a_{m1}b_{pq} & \cdots & \cdots & a_{mn}b_{p1} & a_{mn}b_{p2} & \cdots & a_{mn}b_{pq} \end{bmatrix}.$$

Furthermore, the following useful identities of the Kronecker are presented:

$(\boldsymbol{A} \otimes \boldsymbol{B}) \otimes \boldsymbol{C} = \boldsymbol{A} \otimes (\boldsymbol{B} \otimes \boldsymbol{C})$,

$\boldsymbol{A} \otimes (\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{A} \otimes \boldsymbol{B} + \boldsymbol{A} \otimes \boldsymbol{C}$, where $\boldsymbol{C}$ is a matrix.

$(k\boldsymbol{A}) \otimes \boldsymbol{B} = \boldsymbol{A} \otimes (k\boldsymbol{B}) = k(\boldsymbol{A} \otimes \boldsymbol{B})$, where k is a scalar.

$\boldsymbol{AC} \otimes \boldsymbol{BD} = (\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D})$, denoted as the mixed-product property.

The Kronecker product, $\boldsymbol{A} \otimes \boldsymbol{B}$, is invertible if and only if the matrices involved, $\boldsymbol{A}$ and $\boldsymbol{B}$, are invertible. The inverse is given by:

$(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}$.

In the same way the following properties are given:

$tr(\boldsymbol{A} \otimes \boldsymbol{B}) = tr\boldsymbol{A}tr\boldsymbol{B}$

$(\boldsymbol{A} \otimes \boldsymbol{B})^T = \boldsymbol{A}^T \otimes \boldsymbol{B}^T$.

For further details; refer to Schmidt and Trenkler (2006) and Neudecker and Trenkler (2005a).

Relations Between Vec Operator and Kronecker Product

The following properties of the Kronecker product are used in this work in order to get a convenient representation for several matrix equations.

$(\boldsymbol{B}^\top \otimes \boldsymbol{A})\text{Vec}(\boldsymbol{C}) = \text{Vec}(\boldsymbol{ACB})$

$\text{Vec}(\boldsymbol{ABC}) = (\boldsymbol{I} \otimes \boldsymbol{AB})\text{Vec}(\boldsymbol{C}) = (\boldsymbol{C}^T\boldsymbol{B}^T \otimes \boldsymbol{I})\text{Vec}(\boldsymbol{A})$

$\text{Vec}(\boldsymbol{AB}) = (\boldsymbol{I} \otimes \boldsymbol{A})\text{Vec}(\boldsymbol{B}) = (\boldsymbol{B}^T \otimes \boldsymbol{I})\text{Vec}(\boldsymbol{A})$,

where $\boldsymbol{I}$ is the $n \times n$ identity matrix.

A.4.3 The Commutation Matrix

As mentioned in Magnus and Neudecker (1979) the main properties of the commutation matrix are:

- it transforms the Vec operator form of a matrix into the Vec operator form of its transpose;

  $\boldsymbol{K}_{m,n}$ is the $mn \times mn$ matrix which, for any $m \times n$ matrix $\boldsymbol{A}$, transforms Vec($\boldsymbol{A}$) into Vec($\boldsymbol{A}^T$), i.e.

  $\boldsymbol{K}_{m,n}\ \mathrm{Vec}(\boldsymbol{A}) = \mathrm{Vec}(\boldsymbol{A}^T)$.

- for every $m \times n$ matrix $\boldsymbol{A}$ and every $p \times q$ matrix $\boldsymbol{B}$, it commutes the Kronecker product;

  $\boldsymbol{K}_{p,m}(\boldsymbol{A} \otimes \boldsymbol{B})\boldsymbol{K}_{n,q} = \boldsymbol{B} \otimes \boldsymbol{A}$.

For further details; refer to Magnus and Neudecker (1979).

A.5 Asymptotic Distribution of Quadratic Statistics

A.5.1 Fourth Order Moments of Quadratic Statistics

Consider the independent random vectors $\boldsymbol{z}_i = \boldsymbol{y}_i - \boldsymbol{\mu}$, $i = 1, \ldots, n$, with $\mathrm{E}[\boldsymbol{z}_i] = \boldsymbol{0}$ and $\mathrm{E}[\boldsymbol{z}_i\boldsymbol{z}_i'] = \mathrm{cov}(\boldsymbol{y}_i) = \boldsymbol{\Sigma}$.

Assume that the following third and fourth moment matrices exist:

$$\boldsymbol{\Phi} = \mathrm{E}[\boldsymbol{z}_i \otimes \boldsymbol{z}_i\boldsymbol{z}_i']$$

$$\boldsymbol{\Psi} = \mathrm{E}[\boldsymbol{z}_i\boldsymbol{z}_i' \otimes \boldsymbol{z}_i\boldsymbol{z}_i'].$$

In Kleffe and Rao (1988) the following identity is presented and holds for any general $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}$:

Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be symmetric matrices of order n and $\boldsymbol{a}$ be an n-vector.

Then it follows:

(a) $\mathrm{E}[\boldsymbol{z}_i'\boldsymbol{A}\boldsymbol{z}_i] = tr(\boldsymbol{A}\boldsymbol{\Sigma})$,

(b) $\mathrm{E}[\boldsymbol{a}\boldsymbol{z}_i\boldsymbol{z}_i'\boldsymbol{A}\boldsymbol{z}] = tr(\boldsymbol{a} \otimes \boldsymbol{A})\boldsymbol{\Phi}$,

(c) $\mathrm{E}[\boldsymbol{z}_i'\boldsymbol{A}\boldsymbol{z}_i\boldsymbol{z}_i'\boldsymbol{B}\boldsymbol{z}] = tr(\boldsymbol{A} \otimes \boldsymbol{B})\boldsymbol{\Psi}$.

A.5.2 Third and Fourth Order Moment Matrices of Vec$\boldsymbol{X}'$

Suppose $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$, with $\boldsymbol{x}_i = (X_{i1}, X_{i2})'$ represents a realisation of an i.i.d. bivariate random sample drawn from a bidimensional probability distribution with mean $\mathrm{E}[\boldsymbol{x}_i] = \boldsymbol{\mu}$ and covariance matrix $E[(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})'] = \boldsymbol{\Sigma}$, see Remark 6.1 for more details about the structure of the bivariate random sample $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$.

Similarly to Section 3.1 the random vectors $\boldsymbol{z}_i = \boldsymbol{x}_i - \boldsymbol{\mu}$, $i = 1, \ldots, n$, with existing moments $E[\boldsymbol{z}_i] = \boldsymbol{0}$, $E[\boldsymbol{z}_i\boldsymbol{z}_i'] = \boldsymbol{\Sigma}$, $E[\boldsymbol{z}_i \otimes \boldsymbol{z}_i\boldsymbol{z}_i^{'}] = \boldsymbol{\Phi}$ and $E[\boldsymbol{z}_i\boldsymbol{z}_i^{'} \otimes \boldsymbol{z}_i\boldsymbol{z}_i^{'}] = \boldsymbol{\Psi}$ are considered.

Now, let $\boldsymbol{y} = \mathrm{Vec}(\boldsymbol{X}')$ and $\boldsymbol{z} = \boldsymbol{y} - E[\boldsymbol{y}]$, with $\mathrm{E}[\boldsymbol{y}] = \boldsymbol{1}_n \otimes \boldsymbol{\mu}$ and $\mathrm{cov}(\boldsymbol{y}) = \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}$, from this it follows $\mathrm{E}[\boldsymbol{z}] = \boldsymbol{0}$ and $E[\boldsymbol{z}\boldsymbol{z}'] = \mathrm{cov}(\boldsymbol{y}) = \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}$.

Notice that $\boldsymbol{z}$ decomposes into independent subvectors $\boldsymbol{z}_i$ with dimension $p = 2$, and that the elements of $\boldsymbol{z}_i$ (i.e. $\boldsymbol{z}_i = z_{i1}, z_{i2}$) are **correlated**.

To facilitate notation for the particular bivariate case the following vectors containing random variables are considered, i.e. $\boldsymbol{\mathcal{X}} = (X_1, X_2)'$ and $\boldsymbol{\mathcal{Z}} = (Z_1, Z_2)'$.

The vectors $\boldsymbol{z}_i$, $i = 1, \cdots, n$ are given as follows:

$$\boldsymbol{z}_1 = \begin{pmatrix} z_{11} \\ \vdots \\ z_{1p} \end{pmatrix}, \cdots, \boldsymbol{z}_n = \begin{pmatrix} z_{n1} \\ \vdots \\ z_{np} \end{pmatrix}.$$

As mentioned above the elements of $\boldsymbol{z}_i$ are independent, i.e. $z_{i1}, \cdots, z_{ip}$. Similarly, the vector $\boldsymbol{z}$ contains the following elements:

$$\boldsymbol{z} = \begin{pmatrix} X_{11} - \mu_1 \\ \vdots \\ X_{1p} - \mu_p \\ \vdots \\ \vdots \\ X_{n1} - \mu_1 \\ \vdots \\ X_{np} - \mu_p \end{pmatrix} = \begin{pmatrix} z_{11} \\ \vdots \\ z_{1p} \\ \vdots \\ \vdots \\ z_{n1} \\ \vdots \\ z_{np} \end{pmatrix} = \begin{pmatrix} \boldsymbol{z}_1 \\ \vdots \\ \vdots \\ \boldsymbol{z}_n \end{pmatrix}_{(np\times 1)} .$$

The third and fourth moment matrices $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$, necessary for the computation of the covariance between T_1^{Poly} and T_2^{Poly} as well as the separated variances of those estimators are given as follows:

$$\begin{aligned} \boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{\Phi}_1 \\ \boldsymbol{\Phi}_2 \end{pmatrix} &= \mathrm{E}[\boldsymbol{z}_i \otimes \boldsymbol{z}_i \boldsymbol{z}_i^{'}] = \mathrm{E}[\boldsymbol{\mathcal{Z}} \otimes \boldsymbol{\mathcal{Z}}\boldsymbol{\mathcal{Z}}^{'}] \\ &= \mathrm{E}\left[\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \otimes \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \begin{pmatrix} Z_1 & Z_2 \end{pmatrix} \right], \end{aligned}$$

and

$$\begin{aligned} \boldsymbol{\Psi} = \begin{pmatrix} \boldsymbol{\Psi}_{11} & \boldsymbol{\Psi}_{12} \\ \boldsymbol{\Psi}_{21} & \boldsymbol{\Psi}_{22} \end{pmatrix} &= \mathrm{E}[\boldsymbol{z}_i \boldsymbol{z}_i^{'} \otimes \boldsymbol{z}_i \boldsymbol{z}_i^{'}] = \mathrm{E}[\boldsymbol{\mathcal{Z}}\boldsymbol{\mathcal{Z}}^{'} \otimes \boldsymbol{\mathcal{Z}}\boldsymbol{\mathcal{Z}}^{'}] \\ &= \mathrm{E}\left[\begin{pmatrix} Z_1^2 & Z_1 Z_2 \\ Z_2 Z_1 & Z_2^2 \end{pmatrix} \otimes \begin{pmatrix} Z_1^2 & Z_1 Z_2 \\ Z_2 Z_1 & Z_2^2 \end{pmatrix} \right], \end{aligned}$$

where Φ_j and Ψ_j, $j = 1, 2$ are the third and fourth central moments of the variables Y_j, $j = 1, 2$.

In this work Φ_j and Ψ_j, $j = 1, 2$ has been calculated for selected distributions, based on the relations between non-central and central moments presented in Remark A.3 and the expressions presented in Remark A.6.

In the statistical literature the third and fourth non-central moments of the most commonly used probability distributions have widely been reported, see for example Mood, Graybill and Boes (1974).

In the following remark two useful properties of covariance and correlation are presented. These properties are useful for the numerical calculation of the remaining elements of the matrices showed above.

Remark A.7

Let $\boldsymbol{\mathcal{Z}} = (Z_1, Z_2)'$ *represent a vector containing two random variables with mean* $\mathrm{E}[Z_j] = 0$ *and variance* $\mathrm{var}[Z_j] = \sigma_j^2$, $j = 1, 2$.

From properties of the covariance and correlation, see Mood, Graybill and Boes (1974), it follows:

$$\mathrm{E}[Z_1 Z_2] = \mathrm{E}[Z_1]\mathrm{E}[Z_2] + \mathrm{cov}[Z_1, Z_2].$$

Since the covariance can be obtained from the predefined (known) correlation, i.e. $\mathrm{cov}(Z_1, Z_2) = \mathrm{corr}(Z_1, Z_2)\sigma_1\sigma_2$ *and* $\mathrm{E}[Z_j] = 0$, *then it follows:*

$$\mathrm{E}[Z_1 Z_2] = \mathrm{corr}(Z_1, Z_2)\sigma_{Z_1}\sigma_{Z_2}.$$

Since the standard deviations of Z_j, $j = 1, 2$ *are equal to one, then it follows:*

$$\mathrm{E}[Z_1 Z_2] = \mathrm{corr}(Z_1, Z_2).$$

It can be seen that for uncorrelated random variables, i.e. $\mathrm{corr}(Z_1, Z_2) = 0$, then it follows:

$$E[Z_1 Z_2] = E[Z_1]E[Z_2] = 0.$$

Now, the third and fourth central moment matrices of the vector $\boldsymbol{z}$, with $\boldsymbol{z} = (\boldsymbol{z}_1', \boldsymbol{z}_2', \ldots, \boldsymbol{z}_n')'$ are presented:

$$\boldsymbol{\Phi}_* = \mathrm{E}[\boldsymbol{z} \otimes \boldsymbol{z}\boldsymbol{z}']$$

and

$$\boldsymbol{\Psi}_* = \mathrm{E}[\boldsymbol{z}\boldsymbol{z}' \otimes \boldsymbol{z}\boldsymbol{z}'].$$

In Neudecker and Trenkler (2002) (Theorem 1) the following identities are presented:

(i) $\boldsymbol{\Phi}_* = (\boldsymbol{I}_n \otimes \boldsymbol{K}_{pn} \otimes \boldsymbol{I}_p)(\boldsymbol{G} \otimes \boldsymbol{\Phi})$,

where $\boldsymbol{G} = (\boldsymbol{E}_{11}, \ldots, \boldsymbol{E}_{nn})'$ with $\boldsymbol{E}_{ii} = \boldsymbol{e}_i\boldsymbol{e}_i^{'}$, $\boldsymbol{e}_i$ being the i-th member of the canonical basis of $\mathbb{R}^n$, and commutation matrix $\boldsymbol{K}_{p,n} = \boldsymbol{K}_{pn} = \sum_{i=1}^{p}\sum_{j=1}^{n}(\boldsymbol{H}_{ij} \otimes \boldsymbol{H}_{ij}^{'})$, where $\boldsymbol{H}_{ij}$ is the $p \times n$ matrix with 1 being its (i,j)-th position and zero elsewhere.

Additionally, the following identities have been used:

(a)
$$\boldsymbol{z}_i \otimes \boldsymbol{z}\boldsymbol{z}' = \begin{pmatrix} \boldsymbol{z}_1 \otimes \boldsymbol{z}\boldsymbol{z}' \\ \vdots \\ \boldsymbol{z}_n \otimes \boldsymbol{z}\boldsymbol{z}' \end{pmatrix} = \begin{pmatrix} \underbrace{\begin{pmatrix} z_{11} \\ \vdots \\ z_{1p} \end{pmatrix}}_{p\times 1} \otimes \underbrace{\boldsymbol{z}\boldsymbol{z}'}_{np\times np} \\ \vdots \\ \underbrace{\begin{pmatrix} z_{n1} \\ \vdots \\ z_{np} \end{pmatrix}}_{p\times 1} \otimes \underbrace{\boldsymbol{z}\boldsymbol{z}'}_{np\times np} \end{pmatrix}_{(n^2p^2\times np)}.$$

(b) $\boldsymbol{z}_i \otimes \boldsymbol{z}\boldsymbol{z}' = \boldsymbol{K}_{p,np}(\boldsymbol{z}\boldsymbol{z}' \otimes \boldsymbol{z}_i)$, $\quad i = 1, \ldots, n$. (cf. Magnus and Neudecker, 1979, Theorem 3.1 (viii)),

(c)
$$\boldsymbol{z} \otimes \boldsymbol{z}\boldsymbol{z}' = \begin{pmatrix} \boldsymbol{z}_1 \otimes \boldsymbol{z}\boldsymbol{z}' \\ \vdots \\ \boldsymbol{z}_n \otimes \boldsymbol{z}\boldsymbol{z}' \end{pmatrix} = (\boldsymbol{I}_n \otimes \boldsymbol{K}_{p,np}) \begin{pmatrix} \boldsymbol{z}\boldsymbol{z}' \otimes \boldsymbol{z}_1 \\ \vdots \\ \boldsymbol{z}\boldsymbol{z}' \otimes \boldsymbol{z}_n \end{pmatrix},$$

and

(d)
$$\mathrm{E}[\boldsymbol{z} \otimes \boldsymbol{z}\boldsymbol{z}'] = (\boldsymbol{I}_n \otimes \boldsymbol{K}_{p,np})\mathrm{E} \begin{pmatrix} \boldsymbol{z}\boldsymbol{z}' \otimes \boldsymbol{z}_1 \\ \vdots \\ \boldsymbol{z}\boldsymbol{z}' \otimes \boldsymbol{z}_n \end{pmatrix}.$$

(ii)
$$\boldsymbol{\Psi}_* = (\boldsymbol{I}_{n^2p^2} + \boldsymbol{K}_{np,np})(\boldsymbol{I}_n \otimes \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$$
$$+[\mathrm{Vec}(\boldsymbol{I}_n \otimes \boldsymbol{\Sigma})][\mathrm{Vec}(\boldsymbol{I}_n \otimes \boldsymbol{\Sigma})]'$$

$$+(\boldsymbol{I}_n \otimes \boldsymbol{K}_{pn} \otimes \boldsymbol{I}_p)\{\tilde{\boldsymbol{K}}_{nn} \otimes [\boldsymbol{\Psi} - (\boldsymbol{I}_{p^2} + \boldsymbol{K}_{pp})(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})$$

$$-(\text{Vec}\boldsymbol{\Sigma})(\text{Vec}\boldsymbol{\Sigma})']\}(\boldsymbol{I}_n \otimes \boldsymbol{K}_{np} \otimes \boldsymbol{I}_p), \text{ where } \tilde{\boldsymbol{K}}_{nn} = \sum_{i=1}^{n}(\boldsymbol{E}_{ii} \otimes \boldsymbol{E}_{ii}).$$

Additionally, the following identities have been used:

(a) $\boldsymbol{z}_i\boldsymbol{z}_j^{'} \otimes \boldsymbol{z}\boldsymbol{z}' = \boldsymbol{K}_{p,np}(\boldsymbol{z}\boldsymbol{z}' \otimes \boldsymbol{z}_i\boldsymbol{z}_j^{'})\boldsymbol{K}_{np,p}$,

(b) $\text{E}[\boldsymbol{z}_i\boldsymbol{z}_i^{'} \otimes \boldsymbol{z}_j\boldsymbol{z}_j^{'}] = \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}$,

(c) $\boldsymbol{z}\boldsymbol{z}' \otimes \boldsymbol{z}\boldsymbol{z}' = \begin{pmatrix} \boldsymbol{z}_1\boldsymbol{z}_1^{'} \otimes \boldsymbol{z}\boldsymbol{z}' & \dots & \boldsymbol{z}_1\boldsymbol{z}_n^{'} \otimes \boldsymbol{z}\boldsymbol{z}' \\ \vdots & \dots & \vdots \\ \boldsymbol{z}_n\boldsymbol{z}_1^{'} \otimes \boldsymbol{z}\boldsymbol{z}' & \dots & \boldsymbol{z}_n\boldsymbol{z}_n^{'} \otimes \boldsymbol{z}\boldsymbol{z}' \end{pmatrix}$,

and

(d) $\boldsymbol{z}\boldsymbol{z}' \otimes \boldsymbol{z}\boldsymbol{z}' = (\boldsymbol{I}_n \otimes \boldsymbol{K}_{p,np}) \begin{pmatrix} \boldsymbol{z}\boldsymbol{z}' \otimes \boldsymbol{z}_1\boldsymbol{z}_1^{'} & \dots & \boldsymbol{z}\boldsymbol{z}' \otimes \boldsymbol{z}_1\boldsymbol{z}_n^{'} \\ \vdots & & \vdots \\ \boldsymbol{z}\boldsymbol{z}' \otimes \boldsymbol{z}_n\boldsymbol{z}_1^{'} & \dots & \boldsymbol{z}\boldsymbol{z}' \otimes \boldsymbol{z}_n\boldsymbol{z}_n^{'} \end{pmatrix} (\boldsymbol{I}_n \otimes \boldsymbol{K}_{np,p})$.

A.6 Negative Moments of Random Variables

Calculation of the negative moments of a random variable is a problem that can arise in different situations. Casella and Piegorsch (1985) presented different practical applications (quoted from Mendenhall and Lehman (1960)), where the evaluation of negative moments of random variables is of interest.

They mentioned that the theory behind the existence of negative moments of random variables, such as $\mathrm{E}[X^{-1}]$ or $\mathrm{E}[\bar{x}^{-1}]$, is difficult and not nearly as complete as that involving positive moments. They gave a number of sufficient conditions for the existence of negative moments of a random variable X. In the same field, Casella and Khuri (2002) demonstrated that those conditions can be gathered into one necessary and sufficient condition, which is presented in the following remark.

Remark A.8 (Casella and Khuri (2002), Theorem 4)

Let $f_X(x)$ be a continuous density function defined on $[0,\infty)$. According to Casella and Khuri (2002), Theorem 4 $\mathrm{E}[1/X]$ exists if and only if for any $\epsilon > 0$, there exists a z_0 such that:

$$\int_a^b \frac{f_X(x)}{x}dx < \epsilon, \text{ where } a \text{ and } b \text{ are any two numbers such that } 0 < a < b < z_0.$$

Casella and Piegorsch (1985) pointed out that with the exception of the Cauchy distribution, the existence of at least two positive moments is usually a foregone conclusion. This is not the case, if one is attempting to evaluate negative moments, where non-existence of moments is a much more likely occurrence.

Remark A.9 (Cauchy Distributions)

The Cauchy distribution is an example of a distribution which has no mean, variance or higher moments defined.

For example Lehmann and Shaffer (1988), Theorem 2.1 establishes: a) positive and negative Cauchy tails for the inverted distribution of variables with normal, t, logistic and double–exponential, or uniform distributions on (a,b) with $a < 0 < b$; and b) positive Cauchy tails for many positive random variables, including those with uniform distributions on $(0,a)$ and with exponential distribution, including the χ^2 with 2 degrees of freedom (d.f.).

Some concepts and theorems, useful for understanding the problem of non-existence of negative moments of higher order of random variables, are presented in the following definitions. But in general it is difficult to derive special properties of inverted distributions.

Definition 30 (Cauchy Principal Value)

The Cauchy Principal Value (CPV) of a finite integral of a function f about a point c, with $a \leq c \leq b$ is given by:

$$\int_a^b f(x)dx = \lim_{\epsilon \to 0^+} \left[\int_a^{c-\epsilon} f(x)dx + \int_{c+\epsilon}^b f(x)dx \right].$$

The Cauchy principal value is also known as the principal value integral. For more details refer Whittaker and Watson (1990).

Definition 31 (Gauss Hypergeometric Function)

A generalized hypergeometric function ${}_pF_q(a_1, ..., a_p; b_1, ..., b_q; x)$ is a function which can be defined in the form of a hypergeometric series, i.e., a series for which the ratio of successive terms can be written.
The classical standard hypergeometric series ${}_2F_1$ is given by:

$${}_2F_1(a, b; c; z) = 1 + \frac{(ab)}{(1!c)}z + \frac{(a(a+1)b(b+1))}{(2!c(c+1))}z^2 + ...$$

Definition 32 (Error Function)

The "error function" is a normalized form of the Gaussian function. It is an entire function defined by:

$$\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z \exp[-t^2]dt.$$

The complementary error function, denoted erfc, is defined in terms of the error function:

$$\text{erfc}(z) = 1 - \text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_z^\infty \exp[-t^2]dt.$$

This function can also be defined in terms of an Hypergeometric Function as follows:

$$\text{erfc}(z) = \frac{2z}{\sqrt{\pi}} {}_1F_1(\frac{1}{2}; \frac{3}{2}; -z^2).$$

This functions will be sued in Section B.2.2.

Appendix B

Problem Specific Solutions for the Estimation of NLSS Functions

B.1 Estimation of the Ratio of Means of two Uncorrelated Lognormal Variables

In the following remark different properties of the lognormal distribution are presented.

Remark B.1

Let X be a lognormally distributed random variable and $U = \log(X)$ a normally distributed variable with mean μ and variance σ^2.

Equivalent relationships to obtain μ and σ^2, given the expected value and variance of the lognormally distributed variable X are the following:

$\mu = \log(\mathrm{E}[X]) - \frac{1}{2}\log[1 + \mathrm{var}(X)/(\mathrm{E}[X])^2]$, *and*

$\sigma^2 = \log[1 + \mathrm{var}(X)/(\mathrm{E}[X])^2]$.

In Laurent (1963), it is pointed out that the estimation of the ratio of the expected values of two lognormal variables is usually of interest, i.e.

$$\mathrm{E}[X_2]/\mathrm{E}[X_1] = \exp\left[(\mu_2 - \mu_1) + \frac{(\sigma_2^2 - \sigma_1^2)}{2}\right], \tag{B.1.1}$$

as well as the expected value of their ratio, i.e.

$$\mathrm{E}[X_2/X_1] = \exp\left[(\mu_2 - \mu_1) + \frac{(\sigma_2^2 + \sigma_1^2)}{2}\right], \tag{B.1.2}$$

or the following ratio:

$$\exp[(\mathrm{E}[U_2]]/\exp[(\mathrm{E}[U_1])] = \exp[\mu_2 - \mu_1]. \tag{B.1.3}$$

As pointed out in Laurent (1963), from the viewpoint of descriptive statistics only the third expression is of interest to one who wants to compare the tendencies of two distributions, because the two other expressions may be affected by the influence of the dispersion (or variability) of the variables, represented by σ_1 and σ_2. If $\sigma_1 = \sigma_2$, then $\mathrm{E}[X_2]/\mathrm{E}[X_1] = \exp[(\mathrm{E}[U_2]]/\exp[(\mathrm{E}[U_1])]$.

In Shaban (1981) the estimation of the following general function was considered:

$$\varrho = \exp[a(\mu_2 - \mu_1) + b(\sigma_2^2 - \sigma_1^2)]. \tag{B.1.4}$$

In particular, Equation B.1.4 contains ratios of two lognormally distributed random variables, see Remark B.1. It just depends on how the parameters a and b are chosen.

The ratio of the logarithm variances:

$$\varrho = \exp[\mathrm{var}(\log(X_2))]/\exp[\mathrm{var}(\log(X_1))] = \exp[\sigma_2^2 - \sigma_1^2], \text{ i.e. } a = 0, b = 1.$$

The ratio of the logarithm means:

$$\varrho = \exp\left[\mu_2 + \frac{\sigma_2^2}{2}\right]/\exp\left[\mu_1 + \frac{\sigma_1^2}{2}\right], \text{ i.e. } a = 1, b = 1/2.$$

As mentioned above, Shaban (1981) derived an estimator of ϱ, see Equation B.1.4, with smaller MSE than the ML and the MVUE deduced by Crow (1977), for the following three situations:

1. The shape parameters are equal $\sigma_1^2 = \sigma_2^2 = \sigma^2$ and σ^2 known.
2. The shape parameters are equal $\sigma_1^2 = \sigma_2^2 = \sigma^2$ and σ^2 unknown.
3. $\sigma_1^2 \neq \sigma_2^2$ and both are unknown.

Remark B.2

Assume that in Equation B.1.4 $a = 1$ and $b = 1/2$, then it follows:

(i) *if $\sigma_1^2 = \sigma_2^2$, Equation B.1.4 coincides with Equation B.1.3,*

(ii) *if $\sigma_1^2 \neq \sigma_2^2$, Equation B.1.4 coincides with Equation B.1.1.*

B.1.1 1st Situation: Estimation of the Ratio of Means when the Shape Parameters are Equal and Known

Based on Remark B.2, (i), for this first situation the function to be estimated is given by:

$\varrho = \exp[\mu_2 - \mu_1]$, as presented in Equation B.1.3.

The ML estimator as well as the MVUE, presented by Crow (1977), and the minimal MSE estimators, presented by Shaban (1981), belong to the following class of estimators:

$$\hat{\varrho}_c = \exp[a(\bar{U}_2 - \bar{U}_1)]f(\sigma^2). \tag{B.1.5}$$

Here $\bar{U}_1$ and $\bar{U}_2$ represent the arithmetical mean of the logarithms of the n X_1-variables and n X_2-variables in the sample respectively.

Furthermore, c represents the class of estimators and $f(\sigma^2)$ stands for clarifying that the class of estimators depends on a function of the known parameter σ^2.

Remark B.3

For $a = 1$, it follows that: $\hat{\varrho}_c = \exp[(\bar{U}_2 - \bar{U}_1)]f(\sigma^2)$.

The resulting ML estimator and MVUE for ϱ, as presented in Equation B.1.4, with $a = 1$ are the following:

$$\hat{\varrho}_{ML} = \exp[(\bar{U}_2 - \bar{U}_1)]$$

$$\hat{\varrho}_{MVUE_c} = \exp\left[(\bar{U}_2 - \bar{U}_1) - \frac{\sigma^2}{n}\right].$$

Shaban (1981) pointed out that the MSE of this MVUE is readily seen to be minimum when $f(\sigma^2) = \exp[-3\sigma^2/n]$.

The minimal MSE estimator within the class B.1.5 deduced by Shaban (1981) is given by:

$$\hat{\varrho}_{Shaban} = \exp\left[(\bar{U}_2 - \bar{U}_1) - \frac{3\sigma^2}{n}\right].$$

B.1.2 2nd Situation: Estimation of the Ratio of Means when the Shape Parameters are Equal and Unknown

The function to be estimated is the same as for the first situation, i.e.

$\varrho = \exp[\mu_2 - \mu_1]$, as presented in Equation B.1.3.

The class of estimators considered for ϱ, see Equation B.1.4, when $\sigma_1^2 = \sigma_2^2 = \sigma^2$ and unknown is:

$$\hat{\varrho}_s = \exp[a(\bar{U}_2 - \bar{U}_1)]f(S_p), \tag{B.1.6}$$

where s represents the class of estimators, $S_p = \sum_{i=1}^{n}(U_{2i} - \bar{U}_2)^2 + \sum_{j=1}^{n}(U_{1j} - \bar{U}_1)^2$, and $f(S_p)$ stands for clarifying that this new class of estimators depends on a function of the sample variance S_p.

Both the Maximum Likelihood (ML) estimator and a Minimum Variance Unbiased Estimator (MVUE) of ϱ are members of the class $\hat{\varrho}_s$, presented in Equation B.1.6.

The ML estimator is:

$$\hat{\varrho}_{ML} = \exp[\bar{U}_2 - \bar{U}_1].$$

The MVUE is:

$$\hat{\varrho}_{MVUE_s} = \exp[\bar{U}_2 - \bar{U}_1] g_{2(n-1)} \left(-\frac{(2n-1)}{4n(n-1)^2} S_p \right),$$

with $g_\ell(t) = \sum_{r=0}^{\infty} \frac{\ell^r(\ell+2r)}{\ell(\ell+2)\dots(\ell+2r)} \left(\frac{\ell}{\ell+1}\right)^r \frac{t^r}{r!}$,

this function was introduced by Finney (1941).

Shaban (1981) pointed out that unlike the situation when σ^2 is known it is not possible to find a minimum MSE estimator within the class (B.1.6). However, in the same work the following two methods for obtaining estimators were investigated and their MSEs were studied:

$$\hat{\varrho}_{Shaban1} = \exp[a(\bar{U}_2 - \bar{U}_1) + kS_p/(2n)], \text{ with } k = -3/n$$

$$\hat{\varrho}_{Shaban2} = \exp[a(\bar{U}_2 - \bar{U}_1)] g_{2(n-1)} \left(-\tfrac{3}{2(n-1)^2} \right) S_p,$$

where $g_\ell(.)$ is defined as above.

Shaban (1981) mentioned that especially for the case $\sigma_1^2 = \sigma_2^2 = \sigma^2$, the expressions are too complicated to permit analytical comparison between the MSE of the estimators presented in his work.

However, a numerical comparison was carried out in the aforementioned work, with $\sigma^2 \leq 3$, there it was shown that except in very few cases $\hat{\varrho}_{Shaban1}$ has the smallest MSE in the class ($\hat{\varrho}_s = \exp[(\bar{U}_2 - \bar{U}_1)] f(S_p)$) and $\hat{\varrho}_{Shaban2}$ is shown to be the second best.

Shaban (1981) also pointed out that a limitation of estimators $\hat{\varrho}_{MVUE_s}$ and $\hat{\varrho}_{Shaban2}$ is that the function $g_\ell(.)$ can produce negative values, because their arguments are negative. Due to the exponential term contained in the function $\varrho = \exp[\mu_2 - \mu_1]$, it can be seen that this function cannot take negative values, therefore estimators $\hat{\varrho}_{MVUE_s}$ and $\hat{\varrho}_{Shaban2}$ cannot be considered as reasonable estimators of the function ϱ. Therefore, these estimators will not be taken into account in the simulation study.

B.1.3 3rd Situation: Estimation of the Ratio of Means when the Shape Parameters are Unequal and both Unknown

Based on the second identity presented in Remark B.2, for this third situation the function to be estimated is given by:

$\varrho = \mathrm{E}[X_2]/\mathrm{E}[X_1] = \exp\left[(\mu_2 - \mu_1) + \frac{(\sigma_2^2 - \sigma_1^2)}{2}\right]$, as presented in Equation B.1.1.

When $\sigma_1^2 \neq \sigma_2^2$ and both are unknown, Shaban (1981) considered the following class of estimators:

$$\hat{\varrho}_{ss} = \exp[(\bar{U}_2 - \bar{U}_1)]f(S_1, S_2), \tag{B.1.7}$$

where ss represents the class of estimators and $S_2 = \sum_{i=1}^{n}(U_{2i} - \bar{U}_2)^2$ and $S_1 = \sum_{j=1}^{n}(U_{1j} - \bar{U}_1)^2$. The ML estimator is:

$$\hat{\varrho}_{ML} = \exp\left[(\bar{U}_2 - \bar{U}_1) + (S_2 - S_1)/2n\right].$$

The MVUE is:

$$\hat{\varrho}_{MVUE_{ss}} = \exp[(\bar{U}_2 - \bar{U}_1)]g_{n-1}\left(\frac{-(2nb + a^2)}{2(n-1)^2}S_2\right)g_{n-1}\left(\frac{(2nb - a^2)}{2(n-1)^2}S_1\right).$$

Like the case when $\sigma_1^2 = \sigma_2^2 = \sigma^2$ and σ^2 unknown, it is not possible to find a minimal MSE estimator.

For this case Shaban (1981) also suggested the following two methods and studied their MSEs:

$$\hat{\varrho}_{Shaban1} = \exp[(\bar{U}_2 - \bar{U}_1) + S_w],$$

$$\hat{\varrho}_{Shaban2} = \exp[\bar{U}_2 - \bar{U}_1] g_{n-1}\left(-\frac{(n+3)}{2(n-1)^2}S_2\right) g_{n-1}\left(\frac{(n-3)}{2(n-1)^2}S_1\right),$$

with $S_w = C_2S_2 - C_1S_1$, $C_2 = \dfrac{n^2+3(n+1)}{2n^2}$ and $C_1 = \dfrac{n^2-3(n+1)}{2n^2}$.

In a numerical comparison in Shaban (1981) with $\sigma_1^2 \neq \sigma_2^2$ it was shown that, for all the parameter settings considered, $\hat{\varrho}_{Shaban1}$ was the estimator with the smallest MSE within the class $\hat{\varrho}_{ss} = \exp[\bar{U}_2 - \bar{U}_1] f(S_1, S_2)$.

Additionally, $\hat{\varrho}_{Shaban2}$ and $\hat{\varrho}_{MVUE_{ss}}$ have the limitation that the function $g_\ell(.)$ can produce negative values, because their arguments are also negative. Due to the exponential term contained in the function $\varrho = \exp[a(\mu_2 - \mu_1) + b(\sigma_2^2 - \sigma_1^2)]$, it can be seen that this function cannot take negative values either, then estimators $\hat{\varrho}_{MVUE_{ss}}$ and $\hat{\varrho}_{Shaban2}$ cannot be considered as reasonable estimators of the function ϱ. Therefore, these estimators will not be taken into account in the simulation study and comparisons for this case will only be carried out with $\hat{\varrho}_{ML}$ and $\hat{\varrho}_{Shaban1}$.

B.2 Estimation of the Inverse of the Mean

B.2.1 Estimator Derived by Srivastava and Bhatnagar (1981)

As mentioned in Section 6.3.1, based on the MELO estimator, Srivastava and Bhatnagar (1981) derived a class of estimators, which is free from the limitation of non-existence of moments. They derived exact expressions for the first two moments in the case of normal population. The resulting expressions are merely functions of the sample size n and the ratio $\mathfrak{v} = \sigma^2/\mu^2$, but quite intricate so that no clear inference is drawn. To address this, they proposed large sample approximations.

Based on a random sample of size n and supposing that $\bar{x}$ and $s^2 = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$ are unbiased estimators of the mean μ and variance σ^2 respectively, for the inverse of the mean $f(\mu) = 1/\mu$ they considered the following estimator characterised by a scalar $\mathcal{K}$:

$$T_{Sriv} = \frac{\bar{x}}{\bar{x}^2 + \mathcal{K}(s^2/n)}, \text{ for } \mathcal{K} > 0. \tag{B.2.8}$$

This estimator has moments as long as $\mathcal{K}$ is positive and for $\mathcal{K} = 1$ is asymptotically unbiased (cf. Srivastava and Bhatnagar (1981)).

The exact expression for **normal populations** presented by Srivastava and Bhatnagar (1981) is:

$$T_{Sriv} = \frac{\sqrt{n}}{\sigma} \cdot \frac{z}{z^2 + w} \left[1 - \left(1 - \frac{\mathcal{K}}{n-1}\right) \cdot \frac{w}{z^2 + w}\right]^{-1}, \tag{B.2.9}$$

with $z = \frac{\sqrt{n}\bar{x}}{\sigma}$, $\mathfrak{v} = \frac{\sigma^2}{\mu^2}$ and $w = \frac{(n-1)s^2}{\sigma^2}$.

z follows a normal distribution with mean $(n/\mathfrak{v})^{1/2}$ and variance 1, while w follows a χ^2-distribution with (n-1) degrees of freedom. Further, they are stochastically independent.

The Relative Bias (RB) and Relative Mean Squared Error (RMSE) to order $O(n^{-2})$ of T_{Sriv} are given as follows:

$$\begin{aligned} \text{RB}(T_{Sriv}, f(\mu))_N = \frac{\text{Bias}_{(T_{Sriv}, f(\mu))_N}}{f(\mu)} &= \mu \text{E}[T_{Sriv} - (1/\mu)] \\ &= (1-\mathcal{K})\mathfrak{v}/n + (\mathcal{K}^2 - 6\mathcal{K} + 3)\mathfrak{v}^2/n^2. \end{aligned}$$

$$\begin{aligned} \text{RMSE}(T_{Sriv}, f(\mu))_N = \frac{\text{MSE}_{(T_{Sriv}, f(\mu))_N}}{f(\mu)^2} &= \mu^2 \text{E}[T_{Sriv} - (1/\mu)]^2 \\ &= \mathfrak{v}/n + (\mathcal{K}^2 - 8\mathcal{K} + 9)\mathfrak{v}^2/n^2. \end{aligned}$$

N stands for "normal population".

From the equations above it can be seen that the estimator T_{Sriv} is asymptotically unbiased if $\mathcal{K} = 1$, while it has the smallest MSE, to the order of $O(n^{-2})$, if $\mathcal{K} = 4$.

Srivastava and Bhatnagar (1981), also considered the case of **large-sample approximations for non-normal populations** with finite moments of first, second and third order.

The obtained estimator is given as:

$$T_{Sriv} = \frac{(1+U)}{\mu}[(1+U)^2 + \mathcal{K}(\mathfrak{v}+V)/n]^{-1}, \tag{B.2.10}$$

with $\mathfrak{v} = \dfrac{\sigma^2}{\mu^2}$, $U = \dfrac{(\bar{x}-\mu)}{\mu}$ and $V = \dfrac{(s^2-\sigma^2)}{\mu^2}$.

The Relative Bias (RB) and Relative Mean Squared Error (RMSE) of order $O(n^{-2})$ are given as follows:

$$\begin{aligned}
\mathrm{RB}(T_{Sriv}, f(\mu)) &= (1-\mathcal{K})\mathfrak{v}/n + (\mathcal{K}^2 - 6\mathcal{K} + 3 + (3\mathcal{K}-1)\delta)\mathfrak{v}^2/n^2, \\
\mathrm{RMSE}(T_{Sriv}, f(\mu)) &= \mathfrak{v}/n + (\mathcal{K}^2 - 8\mathcal{K} + 9 + 2(\mathcal{K}-1)\delta)\mathfrak{v}^2/n^2,
\end{aligned}$$

where $\delta = \left(\dfrac{Sk}{\mathfrak{v}}\right)^{1/2}$ and Sk is the Pearson's measure of skewness of the population, given as:

$$Sk = \frac{3(\mu - \tilde{x})}{\sigma}, \text{ with } \tilde{x} \text{ representing the sample median.}$$

If $\delta = 0$, i.e., population is symmetrical, then the expressions for the RB and RMSE are the same as those obtained for normal population.

Srivastava and Bhatnagar (1981) pointed out that the estimator T_{Sriv} has a smaller MSE than $1/\bar{x}$ if $0 < \mathcal{K} < 2(4-\delta)$, for $\delta < 4$. This implies that T_{Sriv} has smaller MSE than $1/\bar{x}$ for all negatively skewed populations and positively skewed populations with $\delta < 4$, provided $\mathcal{K}$ satisfies $0 < \mathcal{K} < 2(4-\delta)$.

In the same paper it is stated that the smallest MSE is achieved if $\mathcal{K} = 2(4-\delta)$.

Srivastava and Bhatnagar (1981) also presented a family of estimators for $f(\mu) = 1/\mu$ on the pattern of T_{Sriv} when σ^2 is known.

This estimator is given as:

$$T_{Sriv*} = \frac{\bar{x}}{\bar{x}^2 + g(\sigma^2/n)},$$

where g is the scalar specifying the estimation and $g > 0$.

In the mentioned work, the expressions for the relative bias and variance of the mentioned estimators are presented.

B.2.2 Unbiased Estimator Deduced by Voinov (1985)

Voinov (1985) also derived unbiased estimators of powers of the inverse of population means, for the following cases:

a) unknown normally-distributed population mean μ and known variance σ^2 .

b) normal population mean $1/\mu^k, k = 1, 2, \ldots$, assuming μ and σ to be unknown. Additionally, $\mu > 0$ is assumed.

The aforementioned estimator for the first case is given by:

$$\hat{g}_k(\bar{x}) = \frac{(-1)^{k-1}\sqrt{n}}{(k-1)!\sqrt{2}} \left(\frac{n}{\sigma^2}\right)^{\frac{k}{2}} \frac{d^{(k-1)}}{dz^{k-1}} \left[\exp\left[\frac{z^2}{2}\right] \operatorname{erfc}\left(\frac{z}{\sqrt{2}}\right)\right]\Bigg|_{z=\frac{\sqrt{n}\bar{x}}{\sigma}} , \qquad \text{(B.2.11)}$$

with erfc given as in Definition 32.

If, for example, $k = 1, 2$, then:

$$T = \frac{1}{\hat{\mu}} = \hat{g}_1(\bar{x}) = \frac{\sqrt{n\pi}}{\sqrt{2}\sigma} \exp\left[\frac{n\bar{x}^2}{2\sigma^2}\right] \operatorname{erfc}\left(\frac{\sqrt{n}\bar{x}}{\sqrt{2}\sigma}\right)$$

and

$$T = \frac{1}{\hat{\mu}^2} = \hat{g}_2(\bar{x}) = \frac{n}{\sigma^2} - \frac{\bar{x}\sqrt{\pi}n^{3/2}}{\sqrt{2}\sigma^3} \exp\left[\frac{n\bar{x}^2}{2\sigma^2}\right] \operatorname{erfc}\left(\frac{\sqrt{n}\bar{x}}{\sqrt{2}\sigma}\right),$$

respectively.

The asymptotic expansions of these equations, if $\frac{\sqrt{n}\mu}{\sqrt{2}\sigma} > 1$, are expressed in the aforementioned work as:

$$\hat{g}_1(\bar{x}) \simeq \frac{1}{\bar{x}} \left[1 + \sum_{m=1}^{\infty} \frac{(-1)^m (2m-1)!!\sigma^{2m}}{n^m \bar{x}^{2m}} \right]$$

and

$$\hat{g}_2(\bar{x}) \simeq \frac{1}{\bar{x}^2} \left[\sum_{m=0}^{\infty} \frac{(-1)^m (2m+1)!!\sigma^{2m}}{n^m \bar{x}^{2m}} \right],$$

respectively. With $(2m-1)!!$ representing the double factorial of $2m-1$, see Definition 18.

The unbiased estimator of $1/\mu^k$ can also be rewritten as:

$$\hat{g}_k(\bar{x}) = -\frac{\sqrt{\pi}(n/\sigma^2)^{k/2}}{\sqrt{2}(k-1)!} \cdot \frac{d^{k-1}}{dz^{k-1}} \left[\exp\left[\frac{z^2}{2}\right] \operatorname{erfc}\left(\frac{z}{\sqrt{2}}\right) \right] \Bigg|_{z=-\frac{\sqrt{n}\bar{x}}{\sigma}}. \tag{B.2.12}$$

In Voinov (1985) it is mentioned that in general the estimator $\hat{g}_k(\bar{x})$ (as presented in Equation B.2.11 and B.2.12) has infinite variance. This divergence is due to the infinity in Expressions B.2.11 and B.2.12, at $\bar{x} = -\infty$ and $\bar{x} = +\infty$, respectively. The variance of B.2.11 and/or B.2.12 will be finite in applications when the probability density function of X is truncated at large $|\bar{x}|$.

In Voinov (1985) an unbiased estimator of powers of the inverse of the normal population mean $\frac{1}{\mu^k}$, $k = 1, 2, \ldots$, assuming μ and σ to be **unknown**, with $\mu > 0$ is also presented. It is given by:

$$\hat{g}_k(\bar{x}, S) = \frac{1}{\bar{x}^k}\, {}_2F_1\left(\frac{k+1}{2}, \frac{k}{2}; \frac{n-1}{2}; -\frac{S(n-1)}{n\bar{x}^2}\right), n > k+1, \ldots$$

It can be seen that this equation represents the sum of the ML estimator $\frac{1}{\bar{x}^k}$, and a correction for its bias.

${}_2F_1(\alpha, \beta; \gamma; z)$ is the Gauss hypergeometric function of the second kind, see Definition 31. Furthermore, $S = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$.

Unfortunately, an unbiased estimator of $\frac{1}{\mu^k}$ if $\mu < 0$, σ^2 being **unknown**, is not obtainable, see Voinov (1985, p. 360) for discussion.

B.2.3 First Moment of ML Estimator According to Voinov (1985)

As mentioned in Section 2.1, Voinov (1985) presented the first moment of the Maximum Likelihood estimator of the inverse population mean, $1/\mu$. This is given by:

$$\mathrm{E}_{\mu}\left(\frac{1}{\bar{x}}\right) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}\frac{1}{x}\exp\left[-\frac{n(x-\mu)^2}{2\sigma^2}\right]dx = \int_{-\infty}^{\infty}\frac{dx}{\varphi(x)}. \tag{B.2.13}$$

Since the first derivative of $\varphi(x)$ exists in the vicinity of $x = 0$ and the second derivative $\varphi''(x)$ also exists ($\varphi'(x)$ being nonzero) it is stated in Voinov (1985) that there exists the Cauchy Principal Value (see Appendix A.6, Definition 30) of the integral in Equation B.2.13. In the same work, the bias of the Maximum Likelihood estimator of the inverse population mean is presented. It is given by:

$$\mathrm{b}\left(\frac{1}{\bar{x}},\frac{1}{\mu}\right) = \mathrm{E}_{\mu}\left(\frac{1}{\bar{x}}\right) - \frac{1}{\mu} = \frac{n}{\sigma^2}\int_{0}^{\mu}\exp\left[\frac{n(t^2-\mu^2)}{2\sigma^2}\right]dt - \frac{1}{\mu}.$$

Appendix C

Additional Proofs

C.1 Lemma A.2

Suppose $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ represents a realisation of an i.i.d. random sample from a p–dimensional distribution with mean $\boldsymbol{\mu} = \mathrm{E}[\boldsymbol{x}_i] = \boldsymbol{g}$ and covariance matrix $\boldsymbol{\Sigma} = \mathrm{E}[(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})']$. Then, consider $\boldsymbol{z}_i = \boldsymbol{x}_i - \boldsymbol{\mu}$, $i = 1, \ldots, n$, with moments given as in Lemma A.1. The elements of $\boldsymbol{z}_i$ are z_{ij}, for $j = 1, \cdots, p$, with $\mathrm{E}[z_{ij}] = 0$, $\mathrm{E}[z_{ij}^2] = \sigma_j^2$, $\mathrm{E}[z_{ij}^3] = \eta_j \sigma_j^3$ and $\mathrm{E}[z_{ij}^4] = \upsilon_j \sigma_j^4$, where $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_n)'$, $\mathrm{Diag}(\boldsymbol{A})$ is a diagonal matrix obtained from $\boldsymbol{A}$ by replacing the off-diagonal elements of $\boldsymbol{A}$ by zero and $\boldsymbol{\Delta}$ is a diagonal matrix with $\upsilon_1 - 3, \cdots, \upsilon_p - 3$ as the diagonal elements.

Additionally, are Ψ and Φ linear functions instead of matrices, Ψ from $\mathbb{R}^p \to \mathcal{M}_{p\times p}$ and Φ from $\mathcal{M}_{p\times p} \to \mathcal{M}_{p\times p}$, where $\mathcal{M}_{p\times p}$ stands for the set of all symmetric $p \times p$ matrices, Φ^* is the conjugated or transposed operator with respect to the usual inner product of matrices.

Furthermore, consider $\boldsymbol{z} = \boldsymbol{y} - \mathrm{E}[\boldsymbol{y}]$, with $\boldsymbol{y} = \mathrm{Vec}\boldsymbol{X}'$ and $\mathrm{E}(\boldsymbol{y}) = \boldsymbol{g}_* = \mathbf{1}_n \otimes \boldsymbol{\mu}$. Then $\mathrm{E}[\boldsymbol{z}] = \mathbf{1}_n \otimes \mathbf{0}$ and $\boldsymbol{V}_* = \mathrm{E}[\boldsymbol{z}\boldsymbol{z}'] = \mathrm{cov}(\boldsymbol{y}) = \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}$.

The following notation will also be introduced:

$$\boldsymbol{\Delta}_* = \boldsymbol{I}_n \otimes \boldsymbol{\Delta}$$

$$\boldsymbol{\eta}_* = \mathbf{1}_n \otimes \boldsymbol{\eta}.$$

Using Remark A.4 and different matrix operations given in Appendix A.4, as well as the following identities:

- $tr\boldsymbol{F}_1 = \frac{1}{n^2} tr(\mathbf{1}_n \mathbf{1}_n' \otimes \boldsymbol{A}) = \frac{1}{n} tr\boldsymbol{A}$
- $tr\boldsymbol{F}_2 = \frac{1}{n} tr(\boldsymbol{I}_n \otimes \boldsymbol{A}) = tr\boldsymbol{A}$

the properties presented in Lemma A.2 are proven as follows:

$$
\begin{aligned}
\text{(i.1) } \Psi(\boldsymbol{F}_1) &= tr(\boldsymbol{F}_1 \boldsymbol{V}_*)\boldsymbol{V}_* + 2\boldsymbol{V}_*\boldsymbol{F}_1\boldsymbol{V}_* \\
&+ \boldsymbol{V}_*\text{Diag}(\boldsymbol{F}_1)\boldsymbol{\Delta}^*\boldsymbol{V}_* \\
&= \tfrac{1}{n^2}[tr((\mathbf{1}_n\mathbf{1}_n' \otimes \boldsymbol{A})(\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}))(\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}) + 2\boldsymbol{V}_*\boldsymbol{F}_1\boldsymbol{V}_* + \boldsymbol{V}_*\text{Diag}(\boldsymbol{F}_1)(\boldsymbol{I}_n \otimes \boldsymbol{\Delta})\boldsymbol{V}_*] \\
&= \tfrac{1}{n}[tr\boldsymbol{A}\boldsymbol{\Sigma}(\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}) + \tfrac{2}{n}\mathbf{1}_n\mathbf{1}_n' \otimes \boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\Sigma} + \tfrac{1}{n}\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}\text{Diag}(\boldsymbol{A})\boldsymbol{\Delta}\boldsymbol{\Sigma}].
\end{aligned}
$$

$$
\begin{aligned}
\text{(ii.1) } \Phi^*(\boldsymbol{F}_1) &= (\boldsymbol{I}_n \otimes \boldsymbol{\Sigma})^{3/2}\text{Diag}(\boldsymbol{F}_1)\boldsymbol{\eta}^* \\
&= (\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}^{3/2})\tfrac{1}{n^2}\boldsymbol{I}_n \otimes \text{Diag}(\boldsymbol{A})(\mathbf{1}_n \otimes \boldsymbol{\eta}) \\
&= [\tfrac{1}{n^2}\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}^{3/2}\text{Diag}(\boldsymbol{A})](\mathbf{1}_n \otimes \boldsymbol{\eta}) \\
&= \tfrac{1}{n^2}\mathbf{1}_n \otimes \boldsymbol{\Sigma}^{3/2}\text{Diag}(\boldsymbol{A})\boldsymbol{\eta}, \\
&\quad \text{with } \text{Diag}(\boldsymbol{F}_1) = \tfrac{1}{n^2}\text{Diag}(\mathbf{1}_n\mathbf{1}_n' \otimes \boldsymbol{A}).
\end{aligned}
$$

Similarly, the expressions for $\boldsymbol{F}_2$ can be demonstrated.

$$
\begin{aligned}
\text{(iii) } \Phi(\boldsymbol{f}) &= \text{Diag}(\boldsymbol{f}\boldsymbol{\eta}_*')\boldsymbol{V}_*^{3/2} \\
&= \tfrac{1}{n}\text{Diag}(\mathbf{1}_n \otimes \boldsymbol{a})(\mathbf{1}_n' \otimes \boldsymbol{\eta}')(\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}^{3/2}) \\
&= \tfrac{1}{n}[\text{Diag}\mathbf{1}_n\mathbf{1}_n' \otimes \text{Diag}(\boldsymbol{a}\boldsymbol{\eta}')](\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}^{3/2}), \text{ since } \text{Diag}(\mathbf{1}_n\mathbf{1}_n') = \boldsymbol{I}_n \\
&= \tfrac{1}{n}\boldsymbol{I}_n \otimes \text{Diag}(\boldsymbol{a}\boldsymbol{\eta}')\boldsymbol{\Sigma}^{3/2}.
\end{aligned}
$$

It is important to point out that the expressions presented in Kleffe and Rao (1988) and Neudecker and Trenkler (2002) are equivalent. As an example it can be mentioned:

(a) $tr(\boldsymbol{F}_1 \otimes \boldsymbol{F}_1)\boldsymbol{\Psi}_* = tr\boldsymbol{F}_1\Psi(\boldsymbol{F}_1)$.

From Neudecker and Trenkler (2002) it is known:

$tr(\boldsymbol{F}_1 \otimes \boldsymbol{F}_1)\boldsymbol{\Psi}_* = \frac{1}{n^3}(tr(\boldsymbol{A} \otimes \boldsymbol{A})\boldsymbol{\Psi} + (n-1)(\alpha + 2\beta))$.

Using the properties introduced by Kleffe and Rao (1988, Section 2.1) for the same examples, the equivalent expression looks like:

$$\begin{aligned}
\text{(iv.1) } tr\boldsymbol{F}_1\Psi(\boldsymbol{F}_1) &= tr(\boldsymbol{F}_1 \otimes \boldsymbol{F}_1)\boldsymbol{\Psi}_* \\
&= tr\boldsymbol{F}_1\left[\tfrac{1}{n}tr\boldsymbol{A\Sigma}(\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}) + \tfrac{2}{n^2}\mathbf{1}_n\mathbf{1}_n' \otimes \boldsymbol{\Sigma A \Sigma} + \tfrac{1}{n^2}\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}\mathrm{Diag}(\boldsymbol{A})\boldsymbol{\Delta\Sigma}\right] \\
&= \tfrac{1}{n^2}\alpha + \tfrac{2}{n^2}\beta + \tfrac{1}{n^3}\boldsymbol{A\Sigma}\mathrm{Diag}(\boldsymbol{A})\boldsymbol{\Delta\Sigma} \\
&= \tfrac{1}{n^2}(\alpha + 2\beta) + \tfrac{1}{n^3}\boldsymbol{A\Sigma}\mathrm{Diag}(\boldsymbol{A})\boldsymbol{\Delta\Sigma} \\
&= \tfrac{(n-1)}{n^3}(\alpha + 2\beta) + \tfrac{1}{n^3}\underbrace{(\alpha + 2\beta + \boldsymbol{A\Sigma}\mathrm{Diag}(\boldsymbol{A})\boldsymbol{\Delta\Sigma})}_{tr\boldsymbol{A}\Psi(\boldsymbol{A})} \\
&= \tfrac{1}{n^3}(tr\boldsymbol{A}\Psi(\boldsymbol{A}) + (n-1)(\alpha + 2\beta)).
\end{aligned}$$

(b)$tr(\boldsymbol{F}_2 \otimes \boldsymbol{F}_2)\boldsymbol{\Psi}_* = tr\boldsymbol{F}_2\Psi(\boldsymbol{F}_2)$.

From Neudecker and Trenkler (2002) it is known:

$tr(\boldsymbol{F}_2 \otimes \boldsymbol{F}_2)\boldsymbol{\Psi}_* = \frac{1}{n}(tr(\boldsymbol{A} \otimes \boldsymbol{A})\boldsymbol{\Psi} + (n-1)\alpha)$.

Using the properties introduced by Kleffe and Rao (1988, Section 2.1) for the same examples, the equivalent expression looks like:

$$\begin{aligned}
\text{(iv.2) } tr\boldsymbol{F}_2\Psi(\boldsymbol{F}_2) &= tr(\boldsymbol{F}_2 \otimes \boldsymbol{F}_2)\boldsymbol{\Psi}_* \\
&= \tfrac{1}{n}tr(\boldsymbol{I}_n \otimes \boldsymbol{A})\left[tr\boldsymbol{A\Sigma}(\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}) + \tfrac{2}{n}\boldsymbol{I}_n' \otimes \boldsymbol{\Sigma A\Sigma} + \tfrac{1}{n}\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}\mathrm{Diag}(\boldsymbol{A})\boldsymbol{\Delta\Sigma}\right] \\
&= \tfrac{1}{n}\underbrace{(\alpha + 2\beta + \boldsymbol{A\Sigma}\mathrm{Diag}(\boldsymbol{A})\boldsymbol{\Delta\Sigma})}_{tr\boldsymbol{A}\Psi(\boldsymbol{A})} + \tfrac{(n-1)}{n}\alpha \\
&= \tfrac{1}{n}(tr\boldsymbol{A}\Psi(\boldsymbol{A}) + (n-1)\alpha).
\end{aligned}$$

$$\begin{aligned}
\text{(v) } tr\boldsymbol{F}_2\Psi(\boldsymbol{F}_1) &= tr(\boldsymbol{F}_2 \otimes \boldsymbol{F}_1)\boldsymbol{\Psi}_* = tr\boldsymbol{F}_1\Psi(\boldsymbol{F}_2) = tr(\boldsymbol{F}_1 \otimes \boldsymbol{F}_2)\boldsymbol{\Psi}_* \\
&= \tfrac{1}{n}tr(\boldsymbol{I}_n \otimes \boldsymbol{A})\left[\tfrac{1}{2}tr\boldsymbol{A\Sigma}(\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}) + \tfrac{2}{n^2}\boldsymbol{\Sigma A\Sigma} + \tfrac{1}{n^2}\mathrm{Diag}(\boldsymbol{A})\boldsymbol{\Delta\Sigma}\right] \\
&= \tfrac{1}{n^2}\underbrace{(\alpha + 2\beta + \boldsymbol{A}\mathrm{Diag}(\boldsymbol{A})\boldsymbol{\Delta\Sigma})}_{tr\boldsymbol{A}\Psi(\boldsymbol{A})} + \tfrac{(n-1)}{n}\alpha \\
&= \tfrac{1}{n^2}(tr\boldsymbol{A}\Psi(\boldsymbol{A}) + (n-1)\alpha).
\end{aligned}$$

$$\text{(vi.1) } \boldsymbol{f}'\Phi^*(\boldsymbol{F}_1) = tr(\boldsymbol{F}_1)\Phi(\boldsymbol{f}) = \tfrac{1}{n^2}tr\boldsymbol{A}\Phi(\boldsymbol{a}) = \tfrac{1}{n^2}\boldsymbol{a}'\Phi^*(\boldsymbol{A}) = \tfrac{1}{n^2}tr(\boldsymbol{a}' \otimes \boldsymbol{A})\boldsymbol{\Phi}.$$

(vi.2) $\boldsymbol{f}'\Phi^*(\boldsymbol{F}_2) \quad = \quad tr(\boldsymbol{F}_2)\Phi(\boldsymbol{f}) = \frac{1}{n}tr\boldsymbol{A}\Phi(\boldsymbol{a}) = \frac{1}{n}tr(\boldsymbol{a}' \otimes \boldsymbol{A})\boldsymbol{\Phi}.$

Identities (vii) - (ix) are obtained by substituting $\boldsymbol{z}_i$ by $\boldsymbol{y}$, $\boldsymbol{a} = \boldsymbol{b}$ by $\boldsymbol{f}$, $\boldsymbol{A}$ by $\boldsymbol{F}_1$, $\boldsymbol{B}$ by $\boldsymbol{F}_2$, $\boldsymbol{\Sigma}$ by $\boldsymbol{V}_* = \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}$ and $\boldsymbol{g}$ by $\boldsymbol{g}_* = \mathbf{1}_n \otimes \boldsymbol{\mu}$ in the equations (vii) - (ix) given in Remark A.4.

■

C.2 Theorem 3.1

Additionally to the identities proved in the last section the following identities will be useful for obtaining the expressions of the variances of T_1^{Poly} and T_2^{Poly} as well as $\text{cov}(T_1^{Poly}, T_2^{Poly})$ introduced in Theorem 3.1.

- $\boldsymbol{V}_*\boldsymbol{F}_1 = \frac{1}{n^2}\mathbf{1}_n\mathbf{1}_n' \otimes \boldsymbol{\Sigma}\boldsymbol{A}$
- $\boldsymbol{V}_*\boldsymbol{F}_2 = \frac{1}{n}\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}\boldsymbol{A}$
- $\boldsymbol{F}_1\boldsymbol{V}_* = \frac{1}{n^2}\mathbf{1}_n\mathbf{1}_n' \otimes \boldsymbol{A}\boldsymbol{\Sigma}$
- $\boldsymbol{F}_2\boldsymbol{V}_* = \frac{1}{n}\boldsymbol{I}_n \otimes \boldsymbol{A}\boldsymbol{\Sigma}$
- $\boldsymbol{F}_1\boldsymbol{V}_*\boldsymbol{F}_1 = \frac{1}{n^3}\mathbf{1}_n\mathbf{1}_n' \otimes \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}$
- $\boldsymbol{F}_2\boldsymbol{V}_*\boldsymbol{F}_2 = \frac{1}{n^2}\boldsymbol{I}_n \otimes \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}$
- $tr\boldsymbol{F}_1\boldsymbol{V}_* = \frac{1}{n}tr\boldsymbol{A}\boldsymbol{\Sigma}$
- $tr\boldsymbol{F}_2\boldsymbol{V}_* = tr\boldsymbol{A}\boldsymbol{\Sigma}$
- $\boldsymbol{V}_*\boldsymbol{F}_1\boldsymbol{g}_* = \frac{1}{n}\mathbf{1}_n \otimes \boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\mu}$
- $\boldsymbol{g}_*\boldsymbol{g}_*'\boldsymbol{F}_1 = \frac{1}{n}\mathbf{1}_n\mathbf{1}_n' \otimes \boldsymbol{\mu}\boldsymbol{\mu}'\boldsymbol{A}$
- $\boldsymbol{g}_*\boldsymbol{g}_*'\boldsymbol{F}_1\boldsymbol{V}_* = \frac{1}{n}\mathbf{1}_n\mathbf{1}_n' \otimes \boldsymbol{\mu}\boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\Sigma}$
- $tr\boldsymbol{F}_1(tr\boldsymbol{F}_1\boldsymbol{V}_*)\boldsymbol{V}_* = \frac{\alpha}{n^2}$
- $tr\boldsymbol{F}_2(tr\boldsymbol{F}_2\boldsymbol{V}_*)\boldsymbol{V}_* = \alpha$
- $tr\boldsymbol{F}_2(tr\boldsymbol{F}_1\boldsymbol{V}_*)\boldsymbol{V}_* = \frac{\alpha}{n}$

The following identities hold for $l = 1, 2$:

- $\boldsymbol{V}_*\boldsymbol{f} = \frac{1}{n}\mathbf{1}_n \otimes \boldsymbol{\Sigma a}$
- $\boldsymbol{g}_*\boldsymbol{f}'_*\boldsymbol{V}_* = \frac{1}{n}\mathbf{1}_n\mathbf{1}'_n \otimes \boldsymbol{\mu a}'\boldsymbol{\Sigma}$
- $\boldsymbol{f}'\boldsymbol{g}_* = \boldsymbol{a}'\boldsymbol{\mu}$
- $\boldsymbol{f}'\boldsymbol{V}_*\boldsymbol{f} = \frac{1}{n}\boldsymbol{a}'\boldsymbol{\Sigma a}$
- $\boldsymbol{F}_l\ \boldsymbol{g}_* = \frac{1}{n^2}(\mathbf{1}_n\mathbf{1}'_n \otimes \boldsymbol{A})(\mathbf{1}_n \otimes \boldsymbol{\mu}) = \frac{1}{n}(\mathbf{1}_n \otimes \boldsymbol{A\mu})$
- $\boldsymbol{g}'_*\boldsymbol{F}_l\boldsymbol{g}_* = \boldsymbol{\mu}'\boldsymbol{A\mu}, l = 1, 2$
- $\boldsymbol{f}'\boldsymbol{V}_*\boldsymbol{F}_l\ \boldsymbol{g}_* = \frac{1}{n}\boldsymbol{a}'\boldsymbol{\Sigma A\mu}$
- $\boldsymbol{g}'_*\boldsymbol{F}_l\ \boldsymbol{V}_*\boldsymbol{F}_l\ \boldsymbol{g}_* = \frac{1}{n}\boldsymbol{\mu}'\boldsymbol{A\Sigma A\mu}$
- $\Phi(\boldsymbol{F}_l\boldsymbol{g}_*) = \frac{1}{n}\Phi(\mathbf{1}_n \otimes \boldsymbol{A\mu}))$,

$$= \frac{1}{n}\mathrm{Diag}[(\mathbf{1}_n \otimes \boldsymbol{A\mu})\boldsymbol{\eta}'_*]\boldsymbol{V}_*^{3/2} = \frac{1}{n}[\mathrm{Diag}(\mathbf{1}_n\mathbf{1}'_n \otimes \boldsymbol{A\mu\eta}')](\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}^{3/2})$$

$$= \frac{1}{n}\boldsymbol{I}_n \otimes \mathrm{Diag}(\boldsymbol{A\mu\eta}')\boldsymbol{\Sigma}^{3/2}$$

In Lemma A.2 the following expressions for $l = 1, 2$ are given:

$$\begin{aligned} \mathrm{var}(T_l^{Poly}) &= \mathrm{var}(\boldsymbol{f}'\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{F}_l\boldsymbol{y}) \\ &= 4\boldsymbol{g}'_*\boldsymbol{F}_l\boldsymbol{V}_*\boldsymbol{F}_l\boldsymbol{g}_* + 4tr(\boldsymbol{g}'_*\boldsymbol{F}_l \otimes \boldsymbol{F}_l)\boldsymbol{\Phi}_* \\ &\quad + tr(\boldsymbol{F}_l \otimes \boldsymbol{F}_l)\boldsymbol{\Psi}_* + 4\boldsymbol{f}'\boldsymbol{V}_*\boldsymbol{F}_l\boldsymbol{g}_* + 2tr(\boldsymbol{f}' \otimes \boldsymbol{F}_l)\boldsymbol{\Phi}_* \\ &\quad + \boldsymbol{f}'\boldsymbol{V}_*\boldsymbol{f} - (tr\boldsymbol{F}_l\boldsymbol{V}_*)^2. \end{aligned}$$

$$\begin{aligned} \mathrm{cov}(T_1^{Poly}, T_2^{Poly}) &= \boldsymbol{f}'[2\boldsymbol{V}_*\boldsymbol{F}_1\boldsymbol{g}_* + \boldsymbol{V}_*\boldsymbol{f} + \Phi^*(\boldsymbol{F}_1)] \\ &\quad + tr\boldsymbol{F}_2[4\boldsymbol{g}_*\boldsymbol{g}'_*\boldsymbol{F}_1\boldsymbol{V}_* + 2\Phi(\boldsymbol{F}_1\boldsymbol{g}_*) \\ &\quad + 2\Phi^*(\boldsymbol{F}_1)\boldsymbol{g}'_* + \Psi(\boldsymbol{F}_1) \\ &\quad + 2\boldsymbol{g}_*\boldsymbol{f}'\boldsymbol{V}_* + \Phi(\boldsymbol{f}) - (tr\boldsymbol{F}_1\boldsymbol{V}_*)\boldsymbol{V}_*]. \end{aligned}$$

The preceding calculations show that the variances of the T_l^{Poly}, for $l = 1, 2$, have the following common terms:

$$4\boldsymbol{g}'_*\boldsymbol{F}_l\boldsymbol{V}_*\boldsymbol{F}_l\boldsymbol{g}_* + 4\boldsymbol{f}'\boldsymbol{V}_*\boldsymbol{F}_l\boldsymbol{g}_* + \boldsymbol{f}'\boldsymbol{V}_*\boldsymbol{f}$$

$$= 4\frac{1}{n}\boldsymbol{\mu}'\boldsymbol{A\Sigma A\mu} + 4\frac{1}{n}\boldsymbol{a}'\boldsymbol{\Sigma A\mu} + \frac{1}{n}\boldsymbol{a}'\boldsymbol{\Sigma a}.$$

This summarises in a constant common term, which is denoted by γ, i.e.

$$\gamma = \tfrac{1}{n}(2\boldsymbol{A\mu} + \boldsymbol{a})'\boldsymbol{\Sigma}(2\boldsymbol{A\mu} + \boldsymbol{a}) \geq 0 \ .$$

It can also be seen that in the expression of the covariance the following constant terms are contained:

$$2\boldsymbol{f}'\boldsymbol{V}_*\boldsymbol{F}_1\boldsymbol{g}_* + \boldsymbol{f}'\boldsymbol{V}_*\boldsymbol{f} + 4tr\boldsymbol{F}_2\boldsymbol{g}_*\boldsymbol{g}_*'\boldsymbol{F}_1\boldsymbol{V}_* + 2tr\boldsymbol{F}_2\boldsymbol{g}_*\boldsymbol{f}'\boldsymbol{V}_*$$

$$= 2\tfrac{1}{n}\boldsymbol{a}'\boldsymbol{\Sigma A\mu} + \tfrac{1}{n}\boldsymbol{a}'\boldsymbol{\Sigma a} + 4\tfrac{1}{n}\boldsymbol{\mu}'\boldsymbol{A\Sigma A\mu} \ + 2\tfrac{1}{n}\boldsymbol{a}'\boldsymbol{\Sigma A\mu}$$

$$= 4\tfrac{1}{n}\boldsymbol{a}'\boldsymbol{\Sigma A\mu} + \tfrac{1}{n}\boldsymbol{a}'\boldsymbol{\Sigma a} + 4\tfrac{1}{n}\boldsymbol{\mu}'\boldsymbol{A\Sigma A\mu},$$

which is also equal to γ.

Then it follows:

$$\begin{aligned}
\mathrm{var}(T_1^{Poly}) &= 4tr(\boldsymbol{g}_*'\boldsymbol{F}_1 \otimes \boldsymbol{F}_1)\boldsymbol{\Phi}_* + tr(\boldsymbol{F}_1 \otimes \boldsymbol{F}_1)\boldsymbol{\Psi}_* \\
&\quad +2tr(\boldsymbol{f}' \otimes \boldsymbol{F}_1)\boldsymbol{\Phi}_* - \tfrac{1}{n^2}(tr\boldsymbol{A\Sigma})^2 + \gamma \\
\mathrm{var}(T_2^{Poly}) &= 4tr(\boldsymbol{g}_*'\boldsymbol{F}_2 \otimes \boldsymbol{F}_2)\boldsymbol{\Phi}_* + tr(\boldsymbol{F}_2 \otimes \boldsymbol{F}_2)\boldsymbol{\Psi}_* \\
&\quad +2tr(\boldsymbol{f}' \otimes \boldsymbol{F}_2)\boldsymbol{\Phi}_* - (tr\boldsymbol{A\Sigma})^2 + \gamma \\
\mathrm{cov}(T_1^{Poly}, T_2^{Poly}) &= \boldsymbol{f}'\Phi^*(\boldsymbol{F}_1) + tr\boldsymbol{F}_2[2\Phi(\boldsymbol{F}_1\boldsymbol{g}_*) \\
&\quad +2\Phi^*(\boldsymbol{F}_1)\boldsymbol{g}_*' + \Psi(\boldsymbol{F}_1) + \Phi(\boldsymbol{f})] - \tfrac{1}{n}(tr\boldsymbol{A\Sigma})^2 + \gamma.
\end{aligned}$$

The following identities hold true:

- $\boldsymbol{F}_1 \otimes \boldsymbol{F}_1 = \frac{1}{n^4}\mathbf{1}_n\mathbf{1}_n' \otimes \boldsymbol{A} \otimes \mathbf{1}_n\mathbf{1}_n' \otimes \boldsymbol{A}$
- $\boldsymbol{F}_2 \otimes \boldsymbol{F}_2 = \frac{1}{n^2}\boldsymbol{I}_n \otimes \boldsymbol{A} \otimes \boldsymbol{I}_n \otimes \boldsymbol{A}$
- $tr(\boldsymbol{F}_1 \otimes \boldsymbol{F}_1)\boldsymbol{\Psi}_* = tr\boldsymbol{F}_1\Psi(\boldsymbol{F}_1) = \frac{1}{n^3}[tr\boldsymbol{A}\Psi(\boldsymbol{A}) + (n-1)(2\beta + \alpha)]$
- $tr(\boldsymbol{F}_2 \otimes \boldsymbol{F}_2)\boldsymbol{\Psi}_* = tr\boldsymbol{F}_2\Psi(\boldsymbol{F}_2) = \frac{1}{n}[tr\boldsymbol{A}\Psi(\boldsymbol{A}) + (n-1)\alpha]$
- $tr(\boldsymbol{F}_2 \otimes \boldsymbol{F}_1)\boldsymbol{\Psi}_* = tr\boldsymbol{F}_1\Psi(\boldsymbol{F}_2) = tr\boldsymbol{F}_2\Psi(\boldsymbol{F}_1) = \frac{1}{n^2}[tr\boldsymbol{A}\Psi(\boldsymbol{A}) + (n-1)\alpha]$
- $tr(\boldsymbol{f}' \otimes \boldsymbol{F}_1)\boldsymbol{\Phi}_* = \boldsymbol{f}'\Phi^*(\boldsymbol{F}_1) = tr(\boldsymbol{F}_1)\Phi(\boldsymbol{f}) = \frac{1}{n^2}\boldsymbol{a}'\Phi^*(\boldsymbol{A}) = \frac{1}{n^2}tr\boldsymbol{A}\Phi(\boldsymbol{a})$
- $tr(\boldsymbol{f}' \otimes \boldsymbol{F}_2)\boldsymbol{\Phi}_* = \boldsymbol{f}'\Phi^*(\boldsymbol{F}_2) = tr(\boldsymbol{F}_2)\Phi(\boldsymbol{f}) = \frac{1}{n}\boldsymbol{a}'\Phi^*(\boldsymbol{A}) = \frac{1}{n}tr\boldsymbol{A}\Phi(\boldsymbol{a})$
- $\Phi^*(\boldsymbol{F}_1)\boldsymbol{g}_*' = \frac{1}{n^2}\mathbf{1}_n\mathbf{1}_n' \otimes \boldsymbol{\Sigma}^{3/2}\mathrm{Diag}(\boldsymbol{A})\boldsymbol{\eta\mu}'$
- $\Phi^*(\boldsymbol{F}_2)\boldsymbol{g}_*' = \frac{1}{n}\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}^{3/2}\mathrm{Diag}(\boldsymbol{A})\boldsymbol{\eta\mu}'$

- $tr(\boldsymbol{g}_*^{'}\boldsymbol{F}_1 \otimes \boldsymbol{F}_1)\boldsymbol{\Phi}_* = tr\boldsymbol{F}_1\Psi^*(\boldsymbol{F}_1)\boldsymbol{g}_*^{'} = \frac{1}{n^2}tr\boldsymbol{A}\Phi^*(\boldsymbol{A})\boldsymbol{\mu}'$
- $tr(\boldsymbol{g}_*^{'}\boldsymbol{F}_2 \otimes \boldsymbol{F}_2)\boldsymbol{\Phi}_* = tr\boldsymbol{F}_2\Psi^*(\boldsymbol{F}_2)\boldsymbol{g}_*^{'} = \frac{1}{n}tr\boldsymbol{A}\Phi^*(\boldsymbol{A})\boldsymbol{\mu}'$
- $tr\boldsymbol{F}_1\Phi(\boldsymbol{F}_1\boldsymbol{g}_*) = tr\boldsymbol{F}_2\Phi^*(\boldsymbol{F}_1)\boldsymbol{g}_*' = \frac{1}{n^2}tr\boldsymbol{A}\Phi^*(\boldsymbol{A})\boldsymbol{\mu}' = \frac{1}{n^2}tr(\boldsymbol{\mu}'\boldsymbol{A}\otimes\boldsymbol{A})\boldsymbol{\Phi}$
- $tr\boldsymbol{F}_2\Phi(\boldsymbol{F}_1\boldsymbol{g}_*) = tr\boldsymbol{F}_2\Phi(\boldsymbol{F}_2\boldsymbol{g}_*) = tr\boldsymbol{F}_2\Psi^*(\boldsymbol{F}_2)\boldsymbol{g}_*^{'} = \frac{1}{n}tr\boldsymbol{A}\Phi^*(\boldsymbol{A})\boldsymbol{\mu}' = \frac{1}{n}tr(\boldsymbol{\mu}'\boldsymbol{A}\otimes\boldsymbol{A})\boldsymbol{\Phi}.$

Form the identities presented above and assuming $\alpha = (tr\boldsymbol{A\Sigma})^2$, then the equations for the variance presented above can be rewritten as:

$$\begin{aligned} \text{var}(T_1^{Poly}) &= 4tr\boldsymbol{F}_1\Psi^*(\boldsymbol{F}_1)\boldsymbol{g}_*^{'} + tr\boldsymbol{F}_1\Psi(\boldsymbol{F}_1) \\ &\quad +2\boldsymbol{f}'\Phi^*(\boldsymbol{F}_1) - \tfrac{1}{n^2}\alpha + \gamma \\ \text{var}(T_2^{Poly}) &= 4tr\boldsymbol{F}_2\Psi^*(\boldsymbol{F}_2)\boldsymbol{g}_*^{'} + tr\boldsymbol{F}_2\Psi(\boldsymbol{F}_2) \\ &\quad +2\boldsymbol{f}'\Phi^*(\boldsymbol{F}_2) - \alpha + \gamma \end{aligned}$$

Now, substituting the identities exposed in Lemma A.2 and those presented above in the corresponding equations of the variances and covariance it follows:

$$\begin{aligned} \text{var}(T_1^{Poly}) &= \tfrac{4}{n^2}tr\boldsymbol{A}\Phi^*(\boldsymbol{A})\boldsymbol{\mu}' + \tfrac{1}{n^3}tr\boldsymbol{A}\Psi(\boldsymbol{A}) \\ &+ \underbrace{\frac{n-1}{n^3}(2\beta-\alpha) + \frac{2}{n^2}tr\boldsymbol{A}\Phi(\boldsymbol{a}) - \frac{\alpha}{n^2}}_{2\frac{(n-1)}{n^3}\beta - \frac{\alpha}{n^3}} + \gamma \\ \text{var}(T_2^{Poly}) &= \tfrac{4}{n}tr\boldsymbol{A}\Phi^*(\boldsymbol{A})\boldsymbol{\mu}' + \tfrac{2}{n}tr\boldsymbol{A}\Phi(\boldsymbol{a}) + \tfrac{1}{n}tr\boldsymbol{A}\Psi(\boldsymbol{A}) + \underbrace{\frac{(n-1)\alpha}{n} - \alpha}_{-\frac{\alpha}{n}} + \gamma \\ \text{cov}(T_1^{Poly}, T_2^{Poly}) &= \tfrac{1}{n^2}tr\boldsymbol{A}\Phi(\boldsymbol{a}) + \tfrac{2}{n}tr\boldsymbol{A}\Phi^*(\boldsymbol{A})\boldsymbol{\mu}' + \tfrac{2}{n^2}tr\boldsymbol{A}\Phi^*(\boldsymbol{A})\boldsymbol{\mu}' \\ &+ \tfrac{1}{n^2}[tr\boldsymbol{A}\Psi(\boldsymbol{A}) + (n-1)\alpha] + \tfrac{1}{n}tr\boldsymbol{A}\Phi(\boldsymbol{a}) - \tfrac{1}{n^2}\alpha + \gamma. \end{aligned}$$

After applying some properties of addition it follows:

(i) $\text{var}(T_1^{Poly}) = \frac{1}{n^2}[4tr\boldsymbol{A}\Phi^*(\boldsymbol{A})\boldsymbol{\mu}' + 2tr\boldsymbol{A}\Phi(\boldsymbol{a}) + \frac{1}{n}tr\boldsymbol{A}\Psi(\boldsymbol{A}) + 2\frac{n-1}{n}\beta - \frac{\alpha}{n}] + \gamma$

(ii) $\text{var}(T_2^{Poly}) = \frac{1}{n}[4tr\boldsymbol{A}\Phi^*(\boldsymbol{A})\boldsymbol{\mu}' + 2tr\boldsymbol{A}\Phi(\boldsymbol{a}) + tr\boldsymbol{A}\Psi(\boldsymbol{A}) - \alpha] + \gamma$

(iii) $\text{cov}(T_1^{Poly}, T_2^{Poly}) = \frac{1}{n^2}[2(n+1)tr\boldsymbol{A}\Phi^*(\boldsymbol{A})\boldsymbol{\mu}' + (n+1)tr\boldsymbol{A}\Phi(\boldsymbol{a}) + tr\boldsymbol{A}\Psi(\boldsymbol{A}) - \alpha] + \gamma.$

■

C.3 Lemma A.3

In Kleffe and Rao (1988, Section 2.1) the following properties are introduced:

(i) $tr(\boldsymbol{A} \otimes \boldsymbol{B})\boldsymbol{\Psi} = tr\boldsymbol{B}\Psi(\boldsymbol{A}) = tr\boldsymbol{A}\Psi(\boldsymbol{B})$

(ii) $tr(\boldsymbol{b}' \otimes \boldsymbol{B})\boldsymbol{\Phi} = tr\boldsymbol{B}\Phi(\boldsymbol{b}) = \boldsymbol{b}'\Phi^*(\boldsymbol{B})$

(iii) $\Psi_{\boldsymbol{z}}(\boldsymbol{A}) = (tr\boldsymbol{A\Sigma})\boldsymbol{\Sigma} + 2\boldsymbol{\Sigma A \Sigma} + \text{Diag}(\boldsymbol{D}_1, \ldots, \boldsymbol{D}_n)$,

with $\boldsymbol{D}_i = \Psi_i(\boldsymbol{A}_{ii}) - 2\boldsymbol{\Sigma}_i\boldsymbol{A}_{ii}\boldsymbol{\Sigma}_i - (tr\boldsymbol{A}_{ii}\boldsymbol{\Sigma}_i)\boldsymbol{\Sigma}_i$.

Since the variance for each $i = 1, \cdots, n$ is the same, $\boldsymbol{\Sigma}_i = \boldsymbol{V}_* = \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}$ and at the same time $\boldsymbol{A}_{ii} = \boldsymbol{A}$, then $\boldsymbol{D}_i = \Psi_u(\boldsymbol{A}) - 2\boldsymbol{\Sigma A \Sigma} - (tr\boldsymbol{A\Sigma})\boldsymbol{\Sigma}$ and $\Psi_u(\boldsymbol{A}) = \sum_{i=1}^{p}\sum_{j=1}^{p} A_{ij}\boldsymbol{\Psi}_{ij}$.

(iv) $\Phi^*_{\boldsymbol{z}}(\boldsymbol{A}) = (\Phi^*_1(\boldsymbol{A}_{11})', \ldots, \Phi^*_n(\boldsymbol{A}_{nn})')'$,

with $\Phi^*_i(\boldsymbol{A}_{ii}) = \Phi^*_u(\boldsymbol{A}) = (tr\boldsymbol{A\Phi}_1, \ldots, tr\boldsymbol{A\Phi}_p)$.

(v) $\Phi_{\boldsymbol{z}}(\boldsymbol{a}) = \text{Diag}(\Phi_1(\boldsymbol{a}_1), \ldots, \Phi_n(\boldsymbol{a}_n))$, with $\Phi_i(\boldsymbol{a}_i) = \Phi_u(\boldsymbol{a}) = \sum_{i=1}^{p} \boldsymbol{a}_i\boldsymbol{\Phi}_i$, with $\boldsymbol{a}_i = \boldsymbol{a}$.

Using Remark A.5 and the aforementioned properties introduced by Kleffe and Rao (1988, Section 2.1), the statements made in Lemma A.3 are proven as follows:

$$
\begin{aligned}
\text{(i.1) } \Psi_{\boldsymbol{z}}(\boldsymbol{F}_1) &= (tr\boldsymbol{F}_1\boldsymbol{V}_*)\boldsymbol{V}_* + 2\boldsymbol{V}_*\boldsymbol{F}_1\boldsymbol{V}_* + \text{Diag}(\boldsymbol{D}_1, \cdots, \boldsymbol{D}_n) \\
&= \Psi_{\boldsymbol{z}}(\boldsymbol{F}_1) = (tr\boldsymbol{F}_1\boldsymbol{V}_*)\boldsymbol{V}_* + 2\boldsymbol{V}_*\boldsymbol{F}_1\boldsymbol{V}_* + \text{Diag}(\boldsymbol{D}_1, \cdots, \boldsymbol{D}_n) \\
&= \tfrac{1}{n^2} tr(\mathbf{1}_n\mathbf{1}_n' \otimes \boldsymbol{A\Sigma})(\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}) + \tfrac{2}{n^2}\mathbf{1}_n\mathbf{1}_n' \otimes \boldsymbol{\Sigma A \Sigma} + \text{Diag}(\boldsymbol{D}_1, \cdots, \boldsymbol{D}_n) \\
&= \tfrac{1}{n} tr(\boldsymbol{A\Sigma})(\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}) + \tfrac{2}{n^2}\mathbf{1}_n\mathbf{1}_n' \otimes \boldsymbol{\Sigma A \Sigma} + \text{Diag}(\boldsymbol{D}_1, \cdots, \boldsymbol{D}_n) \\
&= \tfrac{1}{n}[(tr\boldsymbol{A\Sigma})(\boldsymbol{I}_n \otimes \boldsymbol{\Sigma}) + \tfrac{2}{n}\mathbf{1}_n\mathbf{1}_n' \otimes \boldsymbol{\Sigma A \Sigma} + \tfrac{1}{n}\boldsymbol{I}_n \otimes \text{ Diag}(\boldsymbol{D}_i)].
\end{aligned}
$$

$$
\text{(ii.1)} \quad \Phi^*_{\boldsymbol{z}}(\boldsymbol{F}_1) = \tfrac{1}{n^2}\mathbf{1}_n \otimes \Phi^*_u(\boldsymbol{A})
$$

Similarly, the expressions for $\boldsymbol{F}_2$ can be demonstrated.

$$
\text{(iii)} \quad \Phi_{\boldsymbol{z}}(\boldsymbol{f}) = \tfrac{1}{n}\boldsymbol{I}_n \otimes \text{Diag}(\Phi_u(\boldsymbol{a}))
$$

■

C.4 Theorem 3.2

Additionally to the identities proved in the last section the following identities will be useful for obtaining the expressions of the variances of T_1^{Poly} and T_2^{Poly} as well as $\text{cov}(T_1^{Poly}, T_2^{Poly})$ introduced in Theorem 3.2,

$$\Phi_{\boldsymbol{z}}(\boldsymbol{F}_1\boldsymbol{g}_*) = \tfrac{1}{n}\Phi_{\boldsymbol{z}}((\mathbf{1}_n \otimes \boldsymbol{A\mu})) = \tfrac{1}{n}\boldsymbol{I}_n \otimes \text{ Diag}(\Phi_u(\boldsymbol{A\mu}')) = \Phi_{\boldsymbol{z}}(\boldsymbol{F}_2\boldsymbol{g}_*),$$

with $\boldsymbol{F}_1\boldsymbol{g}_* = \frac{1}{n^2}(\mathbf{1}_n\mathbf{1}_n' \otimes \boldsymbol{A})(\mathbf{1}_n \otimes \boldsymbol{\mu}) = \frac{1}{n}(\mathbf{1}_n \otimes \boldsymbol{A\mu}) = \frac{1}{n}\Phi_{\boldsymbol{z}}(\mathbf{1}_n \otimes \boldsymbol{A\mu})$.

$$\Phi^*_{\boldsymbol{z}}(\boldsymbol{F}_1)\boldsymbol{g}_*' = \tfrac{1}{n^2}\mathbf{1}_n\mathbf{1}_n' \otimes \Phi^*_u(\boldsymbol{A})\boldsymbol{\mu}'$$

$$\Phi^*_{\boldsymbol{z}}(\boldsymbol{F}_2)\boldsymbol{g}_*' = \tfrac{1}{n}\boldsymbol{I}_n \otimes \Phi^*_u(\boldsymbol{A})\boldsymbol{\mu}'$$

$$\begin{aligned} tr(\boldsymbol{g}_*'\boldsymbol{F}_1 \otimes \boldsymbol{F}_2)\boldsymbol{\Phi} &= tr\boldsymbol{F}_2\Phi_u(\boldsymbol{g}_*'\boldsymbol{F}_1) = \boldsymbol{g}_*'\boldsymbol{F}_2\Phi^*_u(\boldsymbol{F}_1) \\ &= tr\left[\tfrac{1}{n^2}(\mathbf{1}_n' \otimes \boldsymbol{\mu}')\mathbf{1}_n\mathbf{1}_n' \otimes \boldsymbol{A} \otimes \tfrac{1}{n}\boldsymbol{I}_n \otimes \boldsymbol{A}\right]\boldsymbol{\Phi} \\ &= \tfrac{1}{n^2}tr(\boldsymbol{\mu}'\boldsymbol{A} \otimes \boldsymbol{A})\boldsymbol{\Phi} = \tfrac{1}{n^2}tr\boldsymbol{A}\Phi_u(\boldsymbol{A\mu}) \\ &= \tfrac{1}{n^2}(\boldsymbol{\mu A})'\Phi^*_u(\boldsymbol{A}), \text{ reminding that } \boldsymbol{A} \text{ is symmetric, i.e. } \boldsymbol{A} = \boldsymbol{A}'. \end{aligned}$$

$$tr(\boldsymbol{g}_*'\boldsymbol{F}_1 \otimes \boldsymbol{F}_2)\boldsymbol{\Phi} = \tfrac{1}{n^2}tr\boldsymbol{A}\Phi(\boldsymbol{A\mu}')$$

$$tr(\boldsymbol{g}_*'\boldsymbol{F}_2 \otimes \boldsymbol{F}_2)\boldsymbol{\Phi} = tr\boldsymbol{F}_2\Phi_u(\boldsymbol{g}_*'\boldsymbol{F}_2) = \boldsymbol{g}_*'\boldsymbol{F}_2\Phi^*_u(\boldsymbol{F}_2) = \tfrac{1}{n}tr\boldsymbol{A}\Phi_u(\boldsymbol{A\mu})$$

$$2tr(\boldsymbol{g}_*'\boldsymbol{F}_1 \otimes \boldsymbol{F}_2)\boldsymbol{\Phi} + tr(\boldsymbol{g}_*'\boldsymbol{F}_2 \otimes \boldsymbol{F}_2)\boldsymbol{\Phi} = \tfrac{2(n+1)}{n}tr\boldsymbol{A}\Phi_u(\boldsymbol{A\mu}) \;.$$

Theorem 3.2 follows directly from the expressions of variance and covariance given in Lemma A.2, the identities given in Lemma A.3 and the additional identities given above.

■

C.5 Lemma A.4

Suppose $x_1, x_2, \ldots, x_n$ represents a realisation of an i.i.d. random sample from a univariate distribution and x_i an arbitrary element of $\boldsymbol{x}$ with $\text{E}[x_i] = \mu$ and variance $\text{var}(x_i) = \sigma^2$. Furthermore, is $\text{E}[\boldsymbol{x}] = \mu\mathbf{1}_n$ and $\text{cov}(\boldsymbol{x}) = \sigma^2\boldsymbol{I}_n$.

Assume that $\boldsymbol{z} = \boldsymbol{x} - \mathrm{E}[\boldsymbol{x}]$, with $\mathrm{E}[\boldsymbol{z}] = \boldsymbol{0}$, $\boldsymbol{g}_* = \mathrm{E}[\boldsymbol{x}] = \mu\boldsymbol{1}_n$ and $\boldsymbol{V}_* = \mathrm{E}[\boldsymbol{z}\boldsymbol{z}'] = \mathrm{cov}(\boldsymbol{x}) = \sigma^2\boldsymbol{I}_n$. Furthermore, assume the elements of $\boldsymbol{z}$ are z_i, with $\mathrm{E}[z_i] = 0$, $\mathrm{E}[z_i^2] = \sigma^2$, $\mathrm{E}[z_i^3] = \eta\sigma^3$ and $\mathrm{E}[z_i^4] = \upsilon\sigma^4$.

Additionally, are Ψ and Φ linear functions instead of matrices, Ψ from $\mathbb{R}^p \to \mathcal{M}_{p\times p}$ and Φ from $\mathcal{M}_{p\times p} \to \mathcal{M}_{p\times p}$, where $\mathcal{M}_{p\times p}$ stands for the set of all symmetric $p \times p$ matrices, Φ^* is the conjugated or transposed operator with respect to the usual inner product of matrices.

The following notation will also be introduced:

$$\boldsymbol{\Delta}_* = \boldsymbol{I}_n\upsilon$$

$$\boldsymbol{\eta}_* = \boldsymbol{1}_n\eta$$

The identities presented in Lemma A.4 were deduced by applying the properties presented in Remark A.4 as well as the following additional properties:

- $tr\boldsymbol{F}_1 = \frac{1}{n^2}trA\ 1_n\boldsymbol{1}_n^{'} = \frac{1}{n}A$
- $tr\boldsymbol{F}_2\boldsymbol{V}_* = tr\frac{A\sigma^2}{n}\boldsymbol{I}_n = A$

The identities presented in Lemma A.4 can be proven as follows:

$$\begin{aligned}
\text{(i.1) } \Psi(\boldsymbol{F}_1) &= \tfrac{1}{n}A\sigma^4\boldsymbol{I}_n + \tfrac{2}{n^2}A\sigma^4\boldsymbol{1}_n\boldsymbol{1}_{n^2}^{'} + \underbrace{\frac{1}{n^2}A\sigma^4\boldsymbol{I}_n\mathrm{Diag}\boldsymbol{1}_n\boldsymbol{1}_n^{'}}_{\frac{A\sigma^4}{n^2}}\boldsymbol{\Delta}\boldsymbol{I}_n \\
&= \tfrac{1}{n}[A\sigma^4\boldsymbol{I}_n + \tfrac{2}{n}A\sigma^4\boldsymbol{1}_n\boldsymbol{1}_n^{'} + \tfrac{1}{n}A\Delta\sigma^4\boldsymbol{I}_n]
\end{aligned}$$

$$\text{(ii.1) } \Phi^*(\boldsymbol{F}_1) = \tfrac{1}{n^2}\underbrace{A\eta\sigma^3\boldsymbol{I}_n\mathrm{Diag}(\boldsymbol{1}_n\boldsymbol{1}_n^{'})\boldsymbol{1}_n}_{\frac{1}{n^2}A\eta\sigma^3\boldsymbol{1}_n}$$

Similarly, the expressions for $\boldsymbol{F}_2$ can be demonstrated.

$$\begin{aligned}
\text{(iii) } \Phi(\boldsymbol{f}) &= \mathrm{Diag}(\boldsymbol{f}\boldsymbol{\eta}_*)\boldsymbol{V}_*^{3/2} = \tfrac{1}{n}\mathrm{Diag}(a\eta\boldsymbol{1}_n\boldsymbol{1}_n')\sigma^3\boldsymbol{I}_n \\
&= \tfrac{1}{n}a\eta\sigma^3\boldsymbol{I}_n, \text{ with } \mathrm{Diag}(\boldsymbol{1}_n\boldsymbol{1}_n') = \boldsymbol{I}_n \text{ and } \boldsymbol{I}_n\boldsymbol{I}_n = \boldsymbol{I}_n \\
&= \text{(iv.1) } tr\boldsymbol{F}_1\Psi(\boldsymbol{F}_1) = tr(\boldsymbol{F}_1 \otimes \boldsymbol{F}_1)\boldsymbol{\Psi}_*, \text{ knowing that } tr\boldsymbol{F}_1 = \tfrac{1}{n}A \\
&= \tfrac{1}{n^3}[tr\boldsymbol{A}\Psi(\boldsymbol{A}) + (n-1)(3\alpha)], \text{ since } \alpha = \beta
\end{aligned}$$

.

(iv.1) $tr\boldsymbol{F}_1\Psi(\boldsymbol{F}_1) = \frac{1}{n^3}[tr\boldsymbol{A}\Psi(\boldsymbol{A}) + (n-1)(3\alpha)]$

(iv.2) $tr\boldsymbol{F}_2\Psi(\boldsymbol{F}_2) \quad = \quad \frac{1}{n}[tr\boldsymbol{A}\Psi(\boldsymbol{A}) + (n-1)\alpha]$

(v) $tr\boldsymbol{F}_2\Psi(\boldsymbol{F}_1) \quad = \quad \frac{1}{n^2}[tr\boldsymbol{A}\Psi(\boldsymbol{A}) + (n-1)\alpha]$, where $tr\boldsymbol{A}\Psi(\boldsymbol{A}) = 3\alpha + \boldsymbol{A\Sigma}Diag\ (\boldsymbol{A})\boldsymbol{\Delta\Sigma}$

(vi.1) $\boldsymbol{f}'\Phi^*(\boldsymbol{F}_1) \quad = \quad tr(\boldsymbol{F}_1)\Phi(\boldsymbol{f}) = \frac{1}{n^2}a\eta\sigma^3 A$

(vi.2) $\boldsymbol{f}'\Phi^*(\boldsymbol{F}_2) \quad = \quad tr(\boldsymbol{F}_2)\Phi(\boldsymbol{f}) = \frac{1}{n}a\eta\sigma^3 A$

■

C.6 Theorem 3.4

Additionally to the identities proved in the last section the following identities will be useful for obtaining the expressions of the variances of T_1^{Poly} and T_2^{Poly} as well as $\text{cov}(T_1^{Poly}, T_2^{Poly})$ introduced in Theorem 3.4.

- $\boldsymbol{V}_*\boldsymbol{F}_1 = \frac{A\sigma^2}{n^2}\mathbf{1}_n\mathbf{1}_n'$
- $\boldsymbol{V}_*\boldsymbol{F}_2 = \frac{A\sigma^2}{n}\boldsymbol{I}_n$
- $\boldsymbol{F}_1\boldsymbol{V}_* = \frac{1}{n^2}\mathbf{1}_n\mathbf{1}_n' \otimes \boldsymbol{A\Sigma}$
- $\boldsymbol{F}_2\boldsymbol{V}_* = \frac{A\sigma^2}{n}\boldsymbol{I}_n$
- $\boldsymbol{F}_3\boldsymbol{V}_* = -\frac{A\sigma^2}{n(n-1)}\boldsymbol{L}$
- $tr\boldsymbol{F}_1\boldsymbol{V}_* = \frac{1}{n^2}trA\sigma^2\mathbf{1}_n\mathbf{1}_n' = \frac{1}{n}A\sigma^2$
- $tr\boldsymbol{F}_2\boldsymbol{V}_* = -tr\frac{A\sigma^2}{n(n-1)}\boldsymbol{L} = tr\frac{A\sigma^2}{n}\boldsymbol{I}_n = A\sigma^2$
- $tr\boldsymbol{F}_3\boldsymbol{V}_* = 0$
- $\boldsymbol{V}_*\boldsymbol{F}_1\boldsymbol{g}_* = \frac{A\mu\sigma^2}{n}\mathbf{1}_n$
- $\boldsymbol{F}_1\boldsymbol{V}_*\boldsymbol{F}_1 = \frac{A^2\sigma^2}{n^3}\mathbf{1}_n\mathbf{1}_n'$

- $\boldsymbol{F}_2\boldsymbol{V}_*\boldsymbol{F}_2 = \frac{A^2\sigma^2}{n^2}\boldsymbol{I}_n$
- $\boldsymbol{F}_3\boldsymbol{V}_*\boldsymbol{F}_3 = \frac{A^2\sigma^2}{n^2(n-1)^2}\left[\boldsymbol{I}_n + (n-2)\,\mathbf{1}_n\mathbf{1}_n^{'}\right]$
- $tr\boldsymbol{F}_1(tr\boldsymbol{F}_1\boldsymbol{V}_*)\boldsymbol{V}_* = \frac{\alpha}{n^2}$
- $tr\boldsymbol{F}_2(tr\boldsymbol{F}_1\boldsymbol{V}_*)\boldsymbol{V}_* = \frac{\alpha}{n}$
- $tr\boldsymbol{F}_2(tr\boldsymbol{F}_2\boldsymbol{V}_*)\boldsymbol{V}_* = \alpha$

The following identities hold true for $l = 1, 2, 3$:

- $\boldsymbol{V}_*\boldsymbol{f} = \frac{a\sigma^2}{n}\mathbf{1}_n$
- $\boldsymbol{g}_*\boldsymbol{f}^{'}\boldsymbol{V}_* = \frac{1}{n}a\mu\sigma^2\mathbf{1}_n\mathbf{1}_n^{'}$
- $\boldsymbol{f}'\boldsymbol{g}_* = \frac{a}{n}\mathbf{1}_n^{'}\mathbf{1}_n\mu = a\mu$
- $\boldsymbol{f}'\boldsymbol{V}_*\boldsymbol{f} = \frac{a}{n}\mathbf{1}_n\frac{a\sigma^2}{n}\mathbf{1}_n = \frac{a^2\sigma}{n}^2.$
- $\boldsymbol{g}_*'\boldsymbol{F}_l\boldsymbol{g}_* = A\mu^2$, $l = 1, 2$
- $\boldsymbol{F}_l\ \boldsymbol{g}_* = \frac{A\mu}{n}\mathbf{1}_n$
- $\boldsymbol{g}_*\boldsymbol{g}_*^{'}\boldsymbol{F}_l = \frac{1}{n}A\mu^2\mathbf{1}_n\mathbf{1}_n^{'}$
- $\boldsymbol{g}_*\boldsymbol{g}_*^{'}\boldsymbol{F}_l\boldsymbol{V}_* = \frac{1}{n}A\mu^2\sigma^2\mathbf{1}_n\mathbf{1}_n^{'}$
- $\boldsymbol{f}'\boldsymbol{V}_*\boldsymbol{F}_l\boldsymbol{g}_* = \frac{aA\mu\sigma^2}{n}$
- $\boldsymbol{g}_*^{'}\boldsymbol{F}_l\boldsymbol{V}_*\boldsymbol{F}_l\boldsymbol{g}_* = \frac{(A\mu\sigma)^2}{n}.$

Furthermore it can be obtained:

- $\boldsymbol{F}_1 \otimes \boldsymbol{F}_1 = \frac{A^2}{n^4}\mathbf{1}_n\mathbf{1}_n^{'} \otimes \mathbf{1}_n\mathbf{1}_n^{'}$
- $\boldsymbol{F}_1 \otimes \boldsymbol{F}_2 = \frac{A^2}{n^3}\mathbf{1}_n\mathbf{1}_n^{'} \otimes \boldsymbol{I}_n$
- $\boldsymbol{F}_2 \otimes \boldsymbol{F}_2 = \frac{A^2}{n^2}\boldsymbol{I}_n \otimes \boldsymbol{I}_n$
- $\boldsymbol{F}_3 \otimes \boldsymbol{F}_3 = \frac{A^2}{n^2(n-1)^2}\boldsymbol{L} \otimes \boldsymbol{L}.$
- $\boldsymbol{g}_*^{'}\boldsymbol{F}_1 \otimes \boldsymbol{F}_1 = \frac{A^2\mu}{n^3}\mathbf{1}_n^{'} \otimes \mathbf{1}_n\mathbf{1}_n^{'}$

- $\boldsymbol{g}_*^{'}\boldsymbol{F}_2 \otimes \boldsymbol{F}_1 = \frac{A^2\mu}{n^2}\mathbf{1}_n^{'} \otimes \mathbf{1}_n\mathbf{1}_n^{'}$
- $\boldsymbol{g}_*^{'}\boldsymbol{F}_2 \otimes \boldsymbol{F}_2 = \frac{A^2\mu}{n^2}\mathbf{1}_n^{'} \otimes \boldsymbol{I}_n$
- $\boldsymbol{g}_*^{'}\boldsymbol{F}_3 \otimes \boldsymbol{F}_3 = -\frac{A^2\mu}{n^2(n-1)^2}\mathbf{1}_n^{'} \otimes \boldsymbol{L}$
- $\boldsymbol{f}' \otimes \boldsymbol{F}_1 = \frac{aA}{n^3}\mathbf{1}_n^{'} \otimes \mathbf{1}_n\mathbf{1}_n^{'}$
- $\boldsymbol{f}' \otimes \boldsymbol{F}_2 = \frac{aA}{n^2}\mathbf{1}_n^{'} \otimes \boldsymbol{I}_n$
- $\boldsymbol{f}' \otimes \boldsymbol{F}_3 = -\frac{aA}{n^2(n-1)}\mathbf{1}_n \otimes \boldsymbol{L}$
- $\Phi(\boldsymbol{F}_1\boldsymbol{g}_*) = \frac{1}{n}\Phi(A\mu\mathbf{1}_n) = \frac{\sigma^2}{n}\underbrace{\mathrm{Diag}(A\mu\mathbf{1}_n\eta\mathbf{1}_n')\boldsymbol{I}_n}_{A\eta\sigma^3\boldsymbol{I}_n} = \frac{1}{n}A\mu\eta\sigma^3\boldsymbol{I}_n$
- $\Phi^*(\boldsymbol{F}_1)\boldsymbol{g}_*' = \frac{A\mu\eta\sigma^3}{n^2}\mathbf{1}_n\mathbf{1}_n'$
- $tr\boldsymbol{F}_1\Phi(\boldsymbol{F}_1\boldsymbol{g}_*) = tr\boldsymbol{F}_2\Phi^*(\boldsymbol{F}_1)\boldsymbol{g}_*' = tr\boldsymbol{F}_1\Phi^*(\boldsymbol{F}_1)\boldsymbol{g}_*' = \frac{A^2\mu\eta\sigma^3}{n^2}$
- $tr\boldsymbol{F}_2\Phi(\boldsymbol{F}_1\boldsymbol{g}_*) = tr\boldsymbol{F}_2\Phi(\boldsymbol{F}_2\boldsymbol{g}_*) = \frac{A^2\mu\eta\sigma^3}{n}$

For the univariate case, $\mathrm{E}[\boldsymbol{x}] = \mu\mathbf{1}_n$ and $\mathrm{cov}(\boldsymbol{x}) = \sigma^2\boldsymbol{I}_n$. Theorem 3.4 follows directly from the expressions of variance and covariance given in Lemma A.2, where the expressions for the multivariate case are presented. Additionally, it is based on Theorem 3.3, the identities presented in Lemma A.4 and those presented above.

■

C.7 Theorem 5.1

Let κ_{min} be given as introduced in Section 5.1, i.e. as:

$$\kappa_{min} = \frac{\mathrm{var}(T_2^{Poly}) - \mathrm{cov}(T_1^{Poly}, T_2^{Poly}) + \iota_2(\iota_2 - \iota_1)}{\mathrm{var}(T_1^{Poly} - T_2^{Poly}) + (\iota_2 - \iota_1)^2}.$$

In Remark 3.1 the properties presented in Lemma A.4 (introduced by Kleffe and Rao (1988)) were used for obtaining the following expressions for the variances and covariances of the estimators:

(i) $\mathrm{var}(T_1^{Poly}) = \frac{1}{n^2}[4tr\boldsymbol{A}\Phi^*(\boldsymbol{A})\boldsymbol{\mu}' + 2tr\boldsymbol{A}\Phi(\boldsymbol{a}) + \frac{1}{n}tr\boldsymbol{A}\Psi(\boldsymbol{A}) + \frac{2(n-1)}{n}\beta - \frac{\alpha}{n}] + \gamma$

(ii) $\text{var}(T_2^{Poly}) = \frac{1}{n}[4tr\boldsymbol{A}\Phi^*(\boldsymbol{A})\boldsymbol{\mu}' + 2tr\boldsymbol{A}\Phi(\boldsymbol{a}) + tr\boldsymbol{A}\Psi(\boldsymbol{A}) - \alpha] + \gamma$

(iii) $\text{var}(T_3^{Poly}) = \dfrac{2}{n(n-1)}\beta + \gamma$

(iv) $\text{cov}(T_1^{Poly}, T_2^{Poly}) = \frac{1}{n^2}[2(n+1)tr\boldsymbol{A}\Phi^*(\boldsymbol{A})\boldsymbol{\mu}' + (n+1)tr\boldsymbol{A}\Phi(\boldsymbol{a}) + tr\boldsymbol{A}\Psi(\boldsymbol{A}) - \alpha] + \gamma.$

The following notation will also be introduced:

$$(\iota_2 - \iota_1) = tr\boldsymbol{A\Sigma} - \tfrac{1}{n}tr\boldsymbol{A\Sigma} = \tfrac{n-1}{n}tr\boldsymbol{A\Sigma}$$

$$\iota_2(\iota_2 - \iota_1) = \tfrac{n-1}{n}(tr\boldsymbol{A\Sigma})^2 = \tfrac{n-1}{n}\alpha$$

$$(\iota_2 - \iota_1)^2 = (\tfrac{n-1}{n})^2\alpha.$$

Using the notation given above and expressions given in Remark A.4 and Lemma 3.1 it follows:

$$\begin{aligned}
&\text{var}(T_2^{Poly}) - \text{cov}(T_1^{Poly}, T_2^{Poly}) \\
&\quad = \tfrac{1}{n^2}[2(n-1)tr\boldsymbol{A}\Phi^*(\boldsymbol{A})\boldsymbol{\mu}' + (n-1)tr\boldsymbol{A}\Phi(\boldsymbol{a}) + (n-1)tr\boldsymbol{A}\Psi(\boldsymbol{A}) + (1-n)\alpha]
\end{aligned}$$

$$\begin{aligned}
&\text{var}(T_2^{Poly}) - \text{cov}(T_1^{Poly}, T_2^{Poly}) + \iota_2(\iota_2 - \iota_1) \\
&\quad = \tfrac{(n-1)}{n^2}[2tr\boldsymbol{A}\Phi^*(\boldsymbol{A})\boldsymbol{\mu}' + tr\boldsymbol{A}\Phi(\boldsymbol{a}) + tr\boldsymbol{A}\Psi(\boldsymbol{A})] + \underbrace{\frac{(1-n)}{n^2}\alpha + \frac{n-1}{n}\alpha}_{(\frac{n-1}{n})^2} \\
&\quad = \tfrac{(n-1)}{n^2}[2tr\boldsymbol{A}\Phi^*(\boldsymbol{A})\boldsymbol{\mu}' + tr\boldsymbol{A}\Phi(\boldsymbol{a}) + tr\boldsymbol{A}\Psi(\boldsymbol{A}) + (n-1)\alpha]
\end{aligned}$$

$$\begin{aligned}
&\text{var}(T_1^{Poly} - T_2^{Poly}) = \text{var}(T_1^{Poly}) + \text{var}(T_2^{Poly}) - 2\text{cov}(T_1^{Poly}, T_2^{Poly}) \\
&\quad = \tfrac{1}{n^2}[\underbrace{\frac{(n^2+1)}{n}tr\boldsymbol{A}\Psi(\boldsymbol{A}) - 2tr\boldsymbol{A}\Psi(\boldsymbol{A})}_{\frac{(n^2-2n+1)}{n}tr\boldsymbol{A}\Psi(\boldsymbol{A})} + \tfrac{(n-1)}{n}2\beta \underbrace{- \frac{(n^2+1)}{n}\alpha + 2\alpha}_{\frac{-(n^2-2n+1)}{n}\alpha}] \\
&\quad = \tfrac{1}{n^2}\left[\tfrac{(n-1)^2}{n}tr\boldsymbol{A}\Psi(\boldsymbol{A}) + \tfrac{(n-1)}{n}2\beta - \tfrac{(n-1)^2}{n}\alpha\right] \\
&\quad = \tfrac{(n-1)}{n^3}\left[(n-1)tr\boldsymbol{A}\Psi(\boldsymbol{A}) + 2\beta - (n-1)\alpha\right]
\end{aligned}$$

$$\begin{aligned}
&\mathrm{var}(T_1^{Poly} - T_2^{Poly}) + (\iota_2 - \iota_1)^2 \\
&\quad = \tfrac{1}{n^3}\left[\tfrac{(n-1)^2}{n} tr\boldsymbol{A}\Psi(\boldsymbol{A}) + \tfrac{(n-1)}{n}2\beta\right] \underbrace{-\frac{(n-1)^2}{n^3}\alpha + (\frac{n-1}{n})^2}_{\frac{(n-1)^2(n-1)}{n^3}\alpha} \\
&\quad = \tfrac{(n-1)}{n^3}[(n-1)tr\boldsymbol{A}\Psi(\boldsymbol{A}) + 2\beta + (n-1)^2\alpha]
\end{aligned}$$

$$\Longrightarrow \kappa_{min} = \frac{n(2tr\boldsymbol{A}\Phi^*(\boldsymbol{A})\boldsymbol{\mu}' + tr\boldsymbol{A}\Phi(\boldsymbol{a}) + tr\boldsymbol{A}\Psi(\boldsymbol{A}) + (n-1)\alpha)}{(n-1)tr\boldsymbol{A}\Psi(\boldsymbol{A}) + 2\beta + (n-1)^2\alpha}.$$

■

C.8 Theorem 5.3

Assume that $\boldsymbol{y} = \mathrm{Vec}\boldsymbol{X}'$ is multinormally distributed with $\mathrm{E}[\boldsymbol{y}] = \boldsymbol{g}_* = \boldsymbol{1}_n \otimes \boldsymbol{\mu}$ and $\mathrm{cov}(\boldsymbol{y}) = \boldsymbol{V}_* = \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}$.

Moreover, let κ_{min} be given as introduced in Section 5.1, i.e. as:

$$\kappa_{min} = \frac{\mathrm{var}(T_2^{Poly}) - \mathrm{cov}(T_1^{Poly}, T_2^{Poly}) + \iota_2(\iota_2 - \iota_1)}{\mathrm{var}(T_1^{Poly} - T_2^{Poly}) + (\iota_2 - \iota_1)^2}.$$

The following notation will also be introduced:

$$(\iota_2 - \iota_1) = tr\boldsymbol{A\Sigma} - \tfrac{1}{n}tr\boldsymbol{A\Sigma} = \tfrac{n-1}{n}tr\boldsymbol{A\Sigma}$$

$$\iota_2(\iota_2 - \iota_1) = \tfrac{n-1}{n}(tr\boldsymbol{A\Sigma})^2 = \tfrac{n-1}{n}\alpha$$

$$(\iota_2 - \iota_1)^2 = (\tfrac{n-1}{n})^2\alpha.$$

Using the notation given above and expressions given in Remark A.4 and Lemma 3.1 it follows:

$$\mathrm{var}(T_2^{Poly}) - \mathrm{cov}(T_1^{Poly}, T_2^{Poly})$$

$= \frac{(n-1)}{n^2} 2\beta$

$\mathrm{var}(T_2^{Poly}) - \mathrm{cov}(T_1^{Poly}, T_2^{Poly}) + \iota_2(\iota_2 - \iota_1)$

$= \frac{(n-1)}{n^2} 2\beta + \frac{n-1}{n}\alpha$

$= \frac{(n-1)}{n^2} [2\beta + n\alpha]$

$\mathrm{var}(T_1^{Poly} - T_2^{Poly}) = \mathrm{var}(T_1^{Poly}) + \mathrm{var}(T_2^{Poly}) - 2\mathrm{cov}(T_1^{Poly}, T_2^{Poly})$

$= \frac{1}{n^2} 2\beta + \gamma + \frac{1}{n} 2\beta + \gamma - \frac{1}{n} 4\beta - 2\gamma$

$= \frac{(n-1)}{n^2} 2\beta$

$\mathrm{var}(T_1^{Poly} - T_2^{Poly}) + (\iota_2 - \iota_1)^2$

$= \frac{(n-1)}{n^2} 2\beta + (\frac{n-1}{n})^2 \alpha$

$= \frac{(n-1)}{n^2} [2\beta + (n-1)\alpha]$

$$\Longrightarrow \kappa_{min} = \frac{\frac{(n-1)}{n^2}[2\beta + n\alpha]}{\frac{(n-1)}{n^2}[2\beta + (n-1)\alpha]} = \frac{2\beta + n\alpha}{2\beta + (n-1)\alpha}.$$

■

C.9 Theorem 5.4

Let $\boldsymbol{z} = \boldsymbol{y} - E[\boldsymbol{y}]$, with $E[\boldsymbol{z}] = \boldsymbol{0}$, $\boldsymbol{V}_* = E[\boldsymbol{z}\boldsymbol{z}'] = cov(\boldsymbol{y}) = \sigma^2 \boldsymbol{I}_n$ and $\boldsymbol{g}_* = \mathrm{E}[\boldsymbol{y}] = \mu \boldsymbol{1}_n$. Moreover, let $\gamma = \frac{\sigma^2}{n}(2A\mu + a)^2$ and $\alpha = \beta = A^2\sigma^4$.

The following notation will also be introduced:

$(\iota_2 - \iota_1) = trA\sigma - \frac{1}{n} trA\sigma = \frac{n-1}{n} A\sigma^2$

$\iota_2(\iota_2 - \iota_1) = \frac{n-1}{n}(A\sigma^2)^2 = \frac{n-1}{n}\alpha$

$(\iota_2 - \iota_1)^2 = (\frac{n-1}{n})^2 \alpha.$

Using the notation given above and the expressions given in Theorem 3.4 it follows:

$$\begin{aligned}
&\mathrm{var}(T_2^{Poly}) - \mathrm{cov}(T_1^{Poly}, T_2^{Poly}) \\
&\quad = \left[\tfrac{4}{n} - \tfrac{2(n+1)}{n^2}\right] A^2\mu\Phi + \left[\tfrac{2}{n} - \tfrac{(n+1)}{n^2}\right] aA\Phi + \left[\tfrac{1}{n} - \tfrac{1}{n^2}\right] A^2\Phi + \left[\tfrac{1}{n^2} - \tfrac{1}{n}\right]\alpha \\
&\quad = \left[\tfrac{2(n-1)}{n^2}\right] A^2\mu\Phi + \left[\tfrac{(n-1)}{n^2}\right] aA\Phi + \left[\tfrac{(n-1)}{n^2}\right] A^2\Phi + \left[\tfrac{(1-n)}{n^2}\right]\alpha
\end{aligned}$$

$$\begin{aligned}
&\mathrm{var}(T_2^{Poly}) - \mathrm{cov}(T_1^{Poly}, T_2^{Poly}) + \iota_2(\iota_2 - \iota_1) \\
&\quad = \left[\tfrac{2(n-1)}{n^2}\right] A^2\mu\Phi + \left[\tfrac{(n-1)}{n^2}\right] aA\Phi + \left[\tfrac{(n-1)}{n^2}\right] A^2\Phi + \left[\tfrac{(n-1)(n-1)}{n^2}\right]\alpha
\end{aligned}$$

$$\begin{aligned}
&\mathrm{var}(T_1^{Poly} - T_2^{Poly}) = \mathrm{var}(T_1^{Poly}) + \mathrm{var}(T_2^{Poly}) - 2\mathrm{cov}(T_1^{Poly}, T_2^{Poly}) \\
&\quad = \underbrace{\left[\frac{(4(n+1) - 4(n+1))}{n^2}\right]}_{0} A^2\mu\Phi + \underbrace{\left[\frac{2(n+1) - 2(n+1)}{n^2}\right]}_{0} aA\Phi \\
&\quad + \left[\tfrac{1}{n^3} + \tfrac{1}{n} - \tfrac{2}{n^2}\right] A^2\Phi + \left[\tfrac{(2n-3)}{n^3} - \tfrac{1}{n} + \tfrac{2}{n}\right]\alpha \\
&\quad = \left[\tfrac{(n-1)(n-1)}{n^3}\right] A^2\Phi + \left[-\tfrac{(n-3)(n-1)}{n^3}\right]\alpha
\end{aligned}$$

$$\begin{aligned}
&\mathrm{var}(T_1^{Poly} - T_2^{Poly}) + (\iota_2 - \iota_1)^2 \\
&\quad = \left[\tfrac{(n-1)^2}{n^3}\right] A^2\Phi + \underbrace{\left[-\frac{(n-3)(n-1)}{n^3} + \frac{(n-1)^2}{n^2}\right]}_{\substack{\frac{(n-1)}{n^3}[n(n-1)-(n-3)] \\ \frac{(n-1)}{n^3}[n^2-2n+3] \\ \frac{(n-1)}{n^3}[2+(n-1)^2]}}\alpha
\end{aligned}$$

$$\Longrightarrow \kappa_{min} = \frac{n[2A^2\mu\Phi + aA\Phi + A^2\Psi + (n-1)\alpha]}{(n-1)A^2\Psi + (2 + (n-1)^2)\alpha}.$$

■

C.10 Theorem 5.5

Assume that $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)'$ represents a realisation of an i.i.d. random sample from a random variable with $\mathrm{E}[x_i] = \mu$ and variance $\mathrm{var}(x_i) = \sigma^2$. Furthermore, is $\mathrm{E}[\boldsymbol{x}] = \mu \mathbf{1}_n$ and $\mathrm{cov}(\boldsymbol{x}) = \sigma^2 \boldsymbol{I}_n$.

Furthermore, assume that $\boldsymbol{x}$ is normally distributed with $\mathrm{E}[\boldsymbol{x}] = \boldsymbol{g}_* = \mu \mathbf{1}_n$ and $\mathrm{cov}(\boldsymbol{x}) = \boldsymbol{V}_* = \sigma^2 \boldsymbol{I}_n$. $\boldsymbol{x} \sim N(\mu \mathbf{1}_n, \sigma^2 \boldsymbol{I}_n)$, with $\Phi = \Phi^N = 0$ and $\Psi = \Psi^N = 3\sigma^2$.

It has already been mentioned that for the univariate case $\alpha = \beta = A^2\sigma^4$, from this equality the expression for κ_{min} for the normal case follows automatically , i.e.

$$\Longrightarrow \kappa_{min} = \frac{2\alpha + n\alpha}{2\alpha + (n-1)\alpha} = \frac{\alpha(2+n)}{\alpha(1+n)} = \frac{(2+n)}{(1+n)}.$$

■

Appendix D

Approximated Results for the Estimation of the Ratio of Means

D.1 MSE($T_l^{Poly}, f^{Poly}(\boldsymbol{\mu})$), $l = 1, 2, 3, \kappa_{min}$ under Different Distribution Assumptions

		$n = 10$				$n = 100$				$n = 1200$			
PS	ρ	T_1^{Poly}	T_2^{Poly}	T_3^{Poly}	$T_{\kappa_{min}}^{Poly}$	T_1^{Poly}	T_2^{Poly}	T_3^{Poly}	$T_{\kappa_{min}}^{Poly}$	T_1^{Poly}	T_2^{Poly}	T_3^{Poly}	$T_{\kappa_{min}}^{Poly}$
1	−0.9	2.41e-06	9.52e-05	1.80e-06	1.77e-06	8.5e-08	2.8e-04	5.743e-08	5.742e-08	7.5e−10	3.6e−04	5.05198e−10	5.05181e−10
	−0.5	3.33e-06	1.22e-04	2.60e-06	2.55e-06	5.4e-08	1.7e-04	3.822e-08	3.822e-08	5.1e−10	2.2e−04	3.55819e−10	3.55813e−10
	0.0	3.43e-06	1.08e-04	2.89e-06	2.82e-06	2.7e-08	6.6e-05	2.060e-08	2.059e-08	2.7e−10	9.8e−05	2.01107e−10	2.01094e−10
	0.5	2.12e-06	5.00e-05	2.00e-06	1.93e-06	9.6e-09	1.4e-05	8.279e-09	8.274e-09	9.9e−11	2.5e−05	8.25935e−11	8.25922e−11
	0.9	4.44e-07	5.67e-06	4.78e-07	4.41e-07	1.4e-09	4.7e-07	1.424e-09	1.420e-09	1.4e−11	1.1e−06	1.36158e−11	1.36146e−11
2	−0.9	9.78e-06	3.79e-04	7.40e-06	7.25e-06	3.3e-07	1.1e-03	2.283e-07	2.283e-07	2.9e−09	1.4e−03	1.97761e−09	1.97760e−09
	−0.5	1.22e-05	3.87e-04	1.03e-05	9.99e-06	1.9e-07	4.7e-04	1.448e-07	1.447e-07	1.7e−09	6.1e−04	1.29588e−09	1.29588e−09
	0.0	1.12e-05	1.93e-04	1.15e-05	1.09e-05	8.6e-08	6.9e-05	8.077e-08	8.068e-08	7.9e−10	9.9e−05	7.28358e−10	7.28353e−10
	0.5	1.17e-05	1.52e-04	1.25e-05	1.16e-05	5.6e-08	2.9e-05	5.427e-08	5.417e-08	4.8e−10	2.3e−05	4.68469e−10	4.68459e−10
	0.9	2.16e-05	7.54e-04	1.73e-05	1.69e-05	7.6e-08	2.2e-04	5.553e-08	5.551e-08	6.6e−10	2.6e−04	4.76146e−10	4.76145e−10
3	−0.9	5.41e-06	2.15e-04	4.02e-06	3.95e-06	1.9e-07	6.6e-04	1.308e-07	1.308e-07	1.7e−09	8.3e−04	1.16191e−09	1.16186e−09
	−0.5	8.29e-06	3.21e-04	6.28e-06	6.16e-06	1.4e-07	4.5e-04	9.478e-08	9.476e-08	1.3e−09	6.1e−04	9.01588e−10	9.01582e−10
	0.0	1.06e-05	3.97e-04	8.24e-06	8.07e-06	8.6e-08	2.7e-04	6.018e-08	6.016e-08	8.7e−10	3.9e−04	6.06545e−10	6.06542e−10
	0.5	1.01e-05	3.70e-04	7.92e-06	7.76e-06	4.7e-08	1.4e-04	3.340e-08	3.340e-08	4.9e−10	2.2e−04	3.49246e−10	3.49241e−10
	0.9	7.28e-06	2.79e-04	5.54e-06	5.44e-06	2.4e-08	7.8e-05	1.683e-08	1.682e-08	2.5e−10	1.1e−04	1.70225e−10	1.70224e−10
4	−0.9	1.52e-05	5.94e-04	1.14e-05	1.12e-05	5.3e-07	1.7e-03	3.574e-07	3.574e-07	4.6e−09	2.2e−03	3.12505e−09	3.12504e−09
	−0.5	1.99e-05	6.96e-04	1.60e-05	1.56e-05	3.2e-07	8.9e-04	2.304e-07	2.303e-07	2.9e−09	1.2e−03	2.11228e−09	2.11227e−09
	0.0	1.87e-05	4.90e-04	1.70e-05	1.64e-05	1.4e-07	2.7e-04	1.198e-07	1.197e-07	1.4e−09	3.9e−04	1.13399e−09	1.13399e−09
	0.5	1.13e-05	1.39e-04	1.23e-05	1.13e-05	5.2e-08	1.5e-05	5.134e-08	5.116e-08	4.9e−10	2.6e−05	4.73399e−10	4.73391e−10
	0.9	6.92e-06	1.68e-04	6.47e-06	6.23e-06	2.5e-08	4.4e-05	2.066e-08	2.065e-08	2.0e−10	4.4e−05	1.70016e−10	1.70015e−10
5	−0.9	7.66e+01	3.06e+03	5.67e+01	5.57e+01	4.1e-01	1.4e+03	2.765e-01	2.765e-01	1.2e−02	5.7e+03	7.95732e−03	7.95730e−03
	−0.5	6.86e+01	2.74e+03	5.08e+01	4.99e+01	4.3e-01	1.5e+03	2.883e-01	2.883e-01	1.1e−02	5.4e+03	7.35634e−03	7.35633e−03
	0.0	7.20e+01	2.88e+03	5.33e+01	5.24e+01	4.8e-01	1.7e+03	3.247e-01	3.246e-01	1.0e−02	4.9e+03	6.79298e−03	6.79297e−03
	0.5	7.10e+01	2.84e+03	5.27e+01	5.17e+01	6.1e-01	2.1e+03	4.090e-01	4.090e-01	9.5e−03	4.6e+03	6.31747e−03	6.31746e−03
	0.9	6.38e+01	2.55e+03	4.73e+01	4.64e+01	9.3e-01	3.2e+03	6.249e-01	6.248e-01	9.0e−03	4.3e+03	6.01541e−03	6.01540e−03
6	−0.9	9.43e+01	3.77e+03	6.99e+01	6.86e+01	4.9e-01	1.6e+03	3.265e-01	3.264e-01	1.4e−02	6.8e+03	9.41201e−03	9.41199e−03
	−0.5	8.19e+01	3.26e+03	6.08e+01	5.97e+01	4.7e-01	1.6e+03	3.143e-01	3.142e-01	1.2e−02	5.9e+03	8.15947e−03	8.15946e−03
	0.0	7.87e+01	3.12e+03	5.86e+01	5.75e+01	4.7e-01	1.6e+03	3.160e-01	3.159e-01	1.0e−02	4.9e+03	6.87412e−03	6.87411e−03
	0.5	6.85e+01	2.72e+03	5.10e+01	5.01e+01	5.2e-01	1.8e+03	3.525e-01	3.524e-01	8.6e−03	4.1e+03	5.74307e−03	5.74306e−03
	0.9	5.37e+01	2.14e+03	3.98e+01	3.91e+01	7.4e-01	2.5e+03	4.961e-01	4.960e-01	7.4e−03	3.6e+03	4.94699e−03	4.94698e−03

Table D.1: MSE($T_l^{Poly}, f^{Poly}(\boldsymbol{\mu})$), $l = 1, 2, 3, \kappa_{min}$ for the Standard Normal Distribution.

		n = 10				n = 100				n = 1200			
PS	ρ	T_1^{Poly}	T_2^{Poly}	T_3^{Poly}	$T_{\kappa_{min}}^{Poly}$	T_1^{Poly}	T_2^{Poly}	T_3^{Poly}	$T_{\kappa_{min}}^{Poly}$	T_1^{Poly}	T_2^{Poly}	T_3^{Poly}	$T_{\kappa_{min}}^{Poly}$
	−0.9	1.76e-05	6.89e-04	1.32e-05	1.30e-05	8.2e-08	2.8e-04	5.553e-08	5.552e-08	7.2e−10	3.4e−04	4.84733e−10	4.84732e−10
	−0.5	8.79e-06	3.37e-04	6.68e-06	6.55e-06	6.7e-08	2.1e-04	4.663e-08	4.662e-08	4.9e−10	2.1e−04	3.45542e−10	3.45541e−10
1	0.0	3.36e-06	1.25e-04	2.61e-06	2.56e-06	4.3e-08	1.2e-04	3.125e-08	3.124e-08	2.7e−10	9.5e−05	2.00691e−10	2.00690e−10
	0.5	1.13e-06	4.62e-05	8.26e-07	8.12e-07	1.9e-08	4.6e-05	1.471e-08	1.470e-08	1.0e−10	2.4e−05	8.56652e−11	8.56649e−11
	0.9	4.19e-07	2.97e-05	1.51e-07	1.50e-07	3.4e-09	8.6e-06	2.545e-09	2.544e-09	1.6e−11	1.3e−06	1.49096e−11	1.49094e−11
	−0.9	7.10e-05	2.69e-03	5.44e-05	5.33e-05	3.05e-07	9.96e-04	2.090e-07	2.098e-07	2.8e−09	1.3e−03	1.91330e−09	1.91330e−09
	−0.5	3.26e-05	1.02e-03	2.77e-05	2.70e-05	2.07e-07	5.37e-04	1.560e-07	1.567e-07	1.7e−09	5.9e−04	1.28380e−09	1.28380e−09
2	0.0	1.21e-05	2.11e-04	1.23e-05	1.16e-05	1.10e-07	1.27e-04	9.911e-08	9.903e-08	8.2e−10	9.6e−05	7.54072e−10	7.54066e−10
	0.5	7.64e-06	1.66e-04	7.39e-06	7.08e-06	6.92e-08	3.08e-05	6.742e-08	6.727e-08	5.4e−10	2.9e−05	5.17796e−10	5.17787e−10
	0.9	9.19e-06	3.66e-04	6.82e-06	6.70e-06	1.04e-07	3.01e-04	7.506e-08	7.504e-08	7.6e−10	3.1e−04	5.47824e−10	5.47823e−10
	−0.9	4.00e-05	1.62e-03	2.94e-05	2.88e-05	1.9e-07	6.6e-04	1.306e-07	1.306e-07	1.7e−09	7.9e−04	1.10881e−09	1.10881e−09
	−0.5	2.20e-05	9.40e-04	1.56e-05	1.53e-05	1.8e-07	6.1e-04	1.236e-07	1.236e-07	1.3e−09	5.9e−04	8.65522e−10	8.65520e−10
3	0.0	1.02e-05	4.67e-04	6.83e-06	6.74e-06	1.5e-07	4.8e-04	9.973e-08	9.971e-08	8.6e−10	3.8e−04	5.95691e−10	5.95690e−10
	0.5	4.43e-06	2.27e-04	2.66e-06	2.63e-06	9.7e-08	3.3e-04	6.620e-08	6.618e-08	5.1e−10	2.2e−04	3.54548e−10	3.54547e−10
	0.9	2.39e-06	1.41e-04	1.22e-06	1.21e-06	5.0e-08	1.8e-04	3.341e-08	3.341e-08	2.7e−10	1.2e−04	1.82211e−10	1.82210e−10
	−0.9	1.10e-04	4.24e-03	8.34e-05	8.18e-05	4.9e-07	1.6e-03	3.387e-07	3.387e-07	4.5e−09	2.1e−03	3.00843e−09	3.00842e−09
	−0.5	5.26e-05	1.87e-03	4.18e-05	4.09e-05	3.7e-07	1.1e-03	2.683e-07	2.682e-07	2.9e−09	1.2e−03	2.06746e−09	2.06746e−09
4	0.0	1.89e-05	5.52e-04	1.65e-05	1.60e-05	2.1e-07	4.9e-04	1.673e-07	1.672e-07	1.4e−09	3.8e−04	1.14836e−09	1.14835e−09
	0.5	7.19e-06	2.07e-04	6.31e-06	6.13e-06	8.3e-08	9.4e-05	7.548e-08	7.542e-08	5.2e−10	2.4e−05	5.08381e−10	5.08370e−10
	0.9	4.75e-06	2.40e-04	2.90e-06	2.87e-06	3.1e-08	6.7e-05	2.525e-08	2.524e-08	2.4e−10	5.7e−05	1.99261e−10	1.99260e−10
	−0.9	6.60e+04	2.91e+06	4.56e+04	4.48e+04	9.2e-01	3.2e+03	6.144e-01	6.143e-01	1.1e−02	5.2e+03	7.00032e−03	7.00031e−03
	−0.5	2.15e+05	1.04e+07	1.38e+05	1.36e+05	1.2	3.9e+03	7.541e-01	7.540e-01	1.0e−02	5.1e+03	6.83225e−03	6.83223e−03
5	0.0	2.78e+05	1.39e+07	1.71e+05	1.69e+05	1.3	4.8e+03	8.855e-01	8.853e-01	1.0e−02	5.0e+03	6.82129e−03	6.82128e−03
	0.5	1.41e+05	6.86e+06	8.96e+04	8.84e+04	1.5	5.3e+03	9.987e-01	9.985e-01	1.1e−02	5.1e+03	6.99229e−03	6.99228e−03
	0.9	3.33e+04	1.49e+06	2.28e+04	2.24e+04	1.6	5.6e+03	1.084	1.084	1.2e−02	5.4e+03	7.47168e−03	7.47166e−03
	−0.9	7.55e+04	3.30e+06	5.25e+04	5.17e+04	1.1	3.7e+03	7.171e-01	7.170e-01	1.2e−02	5.9e+03	8.28306e−03	8.28305e−03
	−0.5	2.29e+05	1.10e+07	1.48e+05	1.46e+05	1.2	4.3e+03	8.189e-01	8.187e-01	1.1e−02	5.4e+03	7.56476e−03	7.56475e−03
6	0.0	2.76e+05	1.38e+07	1.71e+05	1.69e+05	1.3	4.7e+03	8.822e-01	8.821e-01	1.0e−02	4.9e+03	6.86139e−03	6.86138e−03
	0.5	1.31e+05	6.39e+06	8.23e+04	8.13e+04	1.4	4.8e+03	9.040e-01	9.038e-01	9.4e−03	4.5e+03	6.29760e−03	6.29759e−03
	0.9	2.83e+04	1.28e+06	1.91e+04	1.89e+04	1.4	4.7e+03	8.938e-01	8.936e-01	9.2e−03	4.4e+03	6.10536e−03	6.10536e−03

Table D.2: MSE($T_l^{Poly}, f^{Poly}(\boldsymbol{\mu})$), $l = 1, 2, 3, \kappa_{min}$ for the Exponential Distribution.

PS	ρ	$n = 10$: T_1^{Poly}	T_2^{Poly}	T_3^{Poly}	$T_{\kappa_{min}}^{Poly}$	$n = 100$: T_1^{Poly}	T_2^{Poly}	T_3^{Poly}	$T_{\kappa_{min}}^{Poly}$	$n = 1200$: T_1^{Poly}	T_2^{Poly}	T_3^{Poly}	$T_{\kappa_{min}}^{Poly}$
1	−0.9	3.89e-05	1.43e-03	3.03e-05	2.97e-05	1.1e-07	3.7e-04	7.500e-08	7.498e-08	7.3e−10	3.4e−04	4.88381e−10	4.88380e−10
	−0.5	1.62e-05	5.51e-04	1.32e-05	1.29e-05	7.3e-08	2.27e-04	5.148e-08	5.147e-08	4.9e−10	2.1e−04	3.48527e−10	3.48527e−10
	0.0	4.05e-06	1.15e-04	3.58e-06	3.47e-06	3.7e-08	9.8e-05	2.768e-08	2.767e-08	2.7e−10	9.5e−05	2.02572e−10	2.02572e−10
	0.5	5.32e-07	1.07e-05	5.25e-07	5.01e-07	1.3e-08	2.6e-05	1.062e-08	1.062e-08	1.0e−10	2.4e−05	8.64455e−11	8.64451e−11
	0.9	4.40e-08	1.74e-06	3.29e-08	3.23e-08	1.8e-09	1.9e-06	1.590e-09	1.589e-09	1.6e−11	9.3e−07	1.50092e−11	1.50089e−11
2	−0.9	1.58e-04	5.91e-03	1.22e-04	1.20e-04	4.2e-07	1.4e-03	2.899e-07	2.898e-07	2.8e−09	1.3e−03	1.92780e−09	1.92779e−09
	−0.5	6.03e-05	1.81e-03	5.21e-05	5.06e-05	2.4e-07	6.1e-04	1.815e-07	1.815e-07	1.7e−09	6.0e−04	1.29447e−09	1.29447e−09
	0.0	1.48e-05	2.25e-04	1.55e-05	1.45e-05	1.1e-07	1.1e-04	9.395e-08	9.387e-08	8.3e−10	9.6e−05	7.59836e−10	7.59829e−10
	0.5	4.32e-06	7.73e-05	4.38e-06	4.14e-06	5.4e-08	2.0e-05	5.322e-08	5.308e-08	5.4e−10	2.9e−05	5.20867e−10	5.20858e−10
	0.9	1.80e-06	6.95e-05	1.37e-06	1.34e-06	6.9e-08	1.9e-04	4.989e-08	4.988e-08	7.7e−10	3.1e−04	5.51309e−10	5.51308e−10
3	−0.9	8.73e-05	3.19e-03	6.84e-05	6.70e-05	2.6e-07	8.6e-04	1.739e-07	1.739e-07	1.7e−09	7.9e−04	1.11707e−09	1.11706e−09
	−0.5	4.01e-05	1.43e-03	3.18e-05	3.11e-05	1.9e-07	6.3e-04	1.328e-07	1.328e-07	1.4e−09	5.9e−04	8.72908e−10	8.72907e−10
	0.0	1.21e-05	4.16e-04	9.78e-06	9.56e-06	1.2e-07	3.9e-04	8.561e-08	8.559e-08	8.7e−10	3.8e−04	6.01352e−10	6.01351e−10
	0.5	2.10e-06	6.90e-05	1.75e-06	1.70e-06	6.6e-08	2.1e-04	4.641e-08	4.640e-08	5.1e−10	2.2e−04	3.58076e−10	3.58075e−10
	0.9	2.27e-07	7.74e-06	1.85e-07	1.80e-07	2.9e-08	9.6e-05	2.049e-08	2.048e-08	2.7e−10	1.2e−04	1.83868e−10	1.83868e−10
4	−0.9	2.44e-04	9.06e-03	1.90e-04	1.86e-04	6.8e-07	2.2e-03	4.616e-07	4.615e-07	4.5e−09	2.1e−03	3.0312e−09	3.03116e−09
	−0.5	9.72e-05	3.17e-03	8.08e-05	7.88e-05	4.2e-07	1.2e-03	3.018e-07	3.018e-07	2.9e−09	1.2e−03	2.0852e−09	2.08521e−09
	0.0	2.30e-05	5.30e-04	2.18e-05	2.09e-05	1.9e-07	3.9e-04	1.523e-07	1.522e-07	1.4e−09	3.9e−04	1.1586e−09	1.15856e−09
	0.5	3.66e-06	5.34e-05	3.86e-06	3.60e-06	6.0e-08	4.0e-05	5.716e-08	5.708e-08	5.3e−10	2.3e−05	5.1218e−10	5.12162e−10
	0.9	8.50e-07	3.36e-05	6.35e-07	6.23e-07	2.0e-08	3.4e-05	1.720e-08	1.719e-08	2.4e−10	5.6e−05	2.0026e−10	2.00257e−10
5	−0.9	1.17e+02	4.22e+03	9.26e+01	9.07e+01	3.1	1.1e+04	2.084	2.084	1.6e−02	7.6e+03	1.05482e−02	1.05482e−02
	−0.5	4.38e+01	1.62e+03	3.41e+01	3.34e+01	2.5	8.5e+03	1.709	1.709	1.6e−02	7.5e+03	1.04109e−02	1.04108e−02
	0.0	1.52e+01	5.68e+02	1.17e+01	1.15e+01	1.9	6.6e+03	1.319	1.318	1.5e−02	7.3e+03	1.01993e−02	1.01992e−02
	0.5	4.31	1.59e+02	3.35	3.29	1.5	4.9e+03	9.714e-01	9.712e-01	1.5e−02	7.2e+03	9.97703e−03	9.97701e−03
	0.9	1.21	4.32e+01	9.60e-01	9.39e-01	1.1	3.4e+03	6.863e-01	6.861e-01	1.5e−02	7.1e+03	9.78397e−03	9.78396e−03
6	−0.9	1.45e+02	5.24e+03	1.14e+02	1.12e+02	3.7	1.2e+04	2.465	2.465	1.9e−02	8.9e+03	1.24647e−02	1.24646e−02
	−0.5	5.17e+01	1.90e+03	4.04e+01	3.96e+01	2.8	9.5e+03	1.907	1.906	1.7e−02	8.3e+03	1.15275e−02	1.15275e−02
	0.0	1.60e+01	5.90e+02	1.25e+01	1.22e+01	1.9	6.7e+03	1.356	1.355	1.6e−02	7.3e+03	1.02953e−02	1.02952e−02
	0.5	3.83	1.39e+02	3.00	2.94	1.3	4.5e+03	9.043e-01	9.041e-01	1.4e−02	6.5e+03	9.04739e−03	9.04737e−03
	0.9	9.03e-01	3.19e+01	7.22e-01	7.06e-01	8.5e-01	2.9e+03	5.744e-01	5.743e-01	1.2e−02	5.8e+03	8.04104e−03	8.04103e−03

Table D.3: MSE($T_l^{Poly}, f^{Poly}(\boldsymbol{\mu})$), $l = 1, 2, 3, \kappa_{min}$ for the Uniform Distribution.

Appendix E

Simulation Results for the Estimation of the Ratio of Means

E.1 $\widehat{\text{MSE}}(T_l, f(\boldsymbol{\mu}))$, $l = 1, 2, 3, \kappa_{min}$ under Different Distribution Assumptions

		n = 10				n = 100				n = 1200			
PS	ρ	T_1	T_2	T_3	$T_{\kappa_{min}}$	T_1	T_2	T_3	$T_{\kappa_{min}}$	T_1	T_2	T_3	$T_{\kappa_{min}}$
	−0.9	3.886e-03	4.588e-03	3.857e-03	3.824e-03	7.373e-04	1.21e-03	7.358e-04	7.331e-04	3.224e-05	4.2e-04	3.223e-05	3.220e-05
	−0.5	3.065e-03	3.524e-03	3.044e-03	3.020e-03	5.818e-04	8.82e-04	5.808e-04	5.789e-04	2.545e-05	2.7e-04	2.545e-05	2.542e-05
1	0.0	2.040e-03	2.266e-03	2.029e-03	2.015e-03	3.877e-04	5.26e-04	3.872e-04	3.859e-04	1.697e-05	1.2e-04	1.696e-05	1.695e-05
	0.5	1.018e-03	1.090e-03	1.014e-03	1.008e-03	1.937e-04	2.32e-04	1.936e-04	1.929e-04	8.481e-06	3.6e-05	8.481e-06	8.473e-06
	0.9	2.033e-04	2.110e-04	2.028e-04	2.014e-04	3.873e-05	4.13e-05	3.871e-05	3.852e-05	1.696e-06	2.8e-06	1.696e-06	1.694e-06
	−0.9	1.573e-02	1.841e-02	1.561e-02	1.548e-02	2.993e-03	4.79e-03	2.987e-03	2.977e-03	1.307e-04	1.6e-03	1.307e-04	1.305e-04
	−0.5	1.319e-02	1.460e-02	1.312e-02	1.303e-02	2.531e-03	3.40e-03	2.527e-03	2.519e-03	1.102e-04	7.9e-04	1.102e-04	1.101e-04
2	0.0	1.004e-02	1.051e-02	1.002e-02	9.948e-03	1.949e-03	2.14e-03	1.948e-03	1.940e-03	8.455e-05	1.9e-04	8.454e-05	8.447e-05
	0.5	6.936e-03	7.168e-03	6.922e-03	6.885e-03	1.366e-03	1.43e-03	1.365e-03	1.362e-03	5.885e-05	8.6e-05	5.885e-05	5.885e-05
	0.9	4.477e-03	5.047e-03	4.455e-03	4.435e-03	8.962e-04	1.25e-03	8.954e-04	8.946e-04	3.826e-05	3.4e-04	3.826e-05	3.826e-05
	−0.9	8.798e-03	1.042e-02	8.729e-03	8.654e-03	1.667e-03	2.76e-03	1.663e-03	1.657e-03	7.293e-05	9.8e-04	7.291e-05	7.285e-05
	−0.5	7.150e-03	8.385e-03	7.097e-03	7.038e-03	1.354e-03	2.17e-03	1.351e-03	1.347e-03	5.928e-05	7.3e-04	5.927e-05	5.921e-05
3	0.0	5.089e-03	5.899e-03	5.054e-03	5.015e-03	9.653e-04	1.49e-03	9.636e-04	9.604e-04	4.225e-05	4.7e-04	4.224e-05	4.220e-05
	0.5	3.030e-03	3.491e-03	3.010e-03	2.988e-03	5.780e-04	8.75e-04	5.770e-04	5.752e-04	2.523e-05	2.6e-04	2.523e-05	2.521e-05
	0.9	1.382e-03	1.618e-03	1.372e-03	1.363e-03	2.696e-04	4.25e-04	2.692e-04	2.684e-04	1.164e-05	1.4e-04	1.164e-05	1.164e-05
	−0.9	2.433e-02	2.864e-02	2.415e-02	2.395e-02	4.621e-03	7.52e-03	4.612e-03	4.595e-03	2.020e-04	2.6e-03	2.019e-04	2.017e-04
	−0.5	1.937e-02	2.195e-02	1.926e-02	1.911e-02	3.691e-03	5.35e-03	3.685e-03	3.673e-03	1.612e-04	1.4e-03	1.612e-04	1.611e-04
4	0.0	1.320e-02	1.426e-02	1.315e-02	1.306e-02	2.527e-03	3.12e-03	2.524e-03	2.516e-03	1.103e-04	5.5e-04	1.103e-04	1.102e-04
	0.5	7.065e-03	7.315e-03	7.048e-03	6.996e-03	1.362e-03	1.44e-03	1.361e-03	1.354e-03	5.929e-05	8.9e-05	5.929e-05	5.921e-05
	0.9	2.181e-03	2.314e-03	2.175e-03	2.167e-03	4.291e-04	4.94e-04	4.289e-04	4.287e-04	1.849e-05	7.0e-05	1.849e-05	1.849e-05
	−0.9	5.582e+05	–	6.950e+05	2.541e+05	1.027e+02	–	8.966e+02	1.027e+02	1.542	–	2.431	1.542
	−0.5	8.705e+05	–	1.092e+06	5.266e+05	1.086e+02	–	3.448e+02	1.086e+02	1.475	–	2.505	1.474
5	0.0	2.331e+07	–	2.878e+07	3.790e+05	1.189e+02	–	2.660e+02	1.188e+02	1.394	–	8.706	1.393
	0.5	1.722e+06	–	2.139e+06	6.234e+05	1.144e+02	–	2.410e+02	1.142e+02	1.314	–	2.148	1.314
	0.9	1.217e+05	–	1.546e+05	9.057e+04	1.212e+02	–	1.633e+04	1.212e+02	1.250	–	3.372	1.250
	−0.9	6.577e+05	–	8.188e+05	2.957e+05	1.214e+02	–	1.089e+03	1.214e+02	1.826	–	2.832	1.825
	−0.5	1.224e+06	–	1.532e+06	6.925e+05	1.198e+02	–	3.437e+02	1.198e+02	1.646	–	2.406	1.645
6	0.0	1.939e+07	–	2.394e+07	3.326e+05	1.206e+02	–	2.280e+02	1.204e+02	1.425	–	10.94	1.425
	0.5	1.569e+06	–	1.952e+06	6.750e+05	1.045e+02	–	2.424e+02	1.044e+02	1.206	–	2.108	1.206
	0.9	1.039e+05	–	1.310e+05	7.035e+04	9.943e+01	–	1.543e+04	9.943e+01	1.032	–	3.046	1.032

Table E.1: $\widehat{\text{MSE}}(T_l, f(\boldsymbol{\mu}))$, $l = 1, 2, 3, \kappa_{min}$ for the Standard Normal Distribution.

PS	ρ	$n=10$ T_1	T_2	T_3	$T_{\kappa_{min}}$	$n=100$ T_1	T_2	T_3	$T_{\kappa_{min}}$	$n=1200$ T_1	T_2	T_3	$T_{\kappa_{min}}$
1	−0.9	3.793e-03	4.477e-03	3.770e-03	3.756e-03	7.583e-04	1.19e-03	7.574e-04	7.565e-04	3.137e-05	3.9e-04	3.137e-05	3.135e-05
	−0.5	2.967e-03	3.079e-03	2.982e-03	2.955e-03	5.976e-04	7.68e-04	5.981e-04	5.969e-04	2.476e-05	2.1e-04	2.476e-05	2.476e-05
	0.0	1.964e-03	1.849e-03	1.988e-03	1.848e-03	3.978e-04	4.30e-04	3.988e-04	3.912e-04	1.651e-05	8.9e-05	1.651e-05	1.650e-05
	0.5	9.757e-04	8.524e-04	9.932e-04	7.826e-04	1.987e-04	1.85e-04	1.994e-04	1.845e-04	8.252e-06	2.4e-05	8.253e-06	8.234e-06
	0.9	1.942e-04	1.611e-04	1.984e-04	1.101e-04	3.970e-05	3.29e-05	3.987e-05	2.817e-05	1.650e-06	2.0e-06	1.650e-06	1.612e-06
2	−0.9	1.555e-02	1.912e-02	1.537e-02	1.507e-02	3.084e-03	4.97e-03	3.077e-03	3.059e-03	1.271e-04	1.5e-03	1.271e-04	1.269e-04
	−0.5	1.316e-02	1.453e-02	1.309e-02	1.298e-02	2.608e-03	3.33e-03	2.606e-03	2.599e-03	1.073e-04	6.6e-04	1.073e-04	1.072e-04
	0.0	1.010e-02	1.016e-02	1.012e-02	1.009e-02	2.006e-03	2.07e-03	2.007e-03	2.004e-03	8.264e-05	1.5e-04	8.264e-05	8.264e-05
	0.5	7.011e-03	6.620e-03	7.064e-03	6.297e-03	1.400e-03	1.32e-03	1.402e-03	1.317e-03	5.807e-05	7.9e-05	5.807e-05	5.786e-05
	0.9	4.482e-03	4.189e-03	4.546e-03	4.189e-03	9.090e-04	1.02e-03	9.117e-04	8.918e-04	3.849e-05	2.7e-04	3.850e-05	3.848e-05
3	−0.9	8.512e-03	9.743e-03	8.495e-03	8.494e-03	1.712e-03	2.62e-03	1.711e-03	1.711e-03	7.104e-05	8.9e-04	7.104e-05	7.102e-05
	−0.5	6.787e-03	6.681e-03	6.869e-03	6.585e-03	1.385e-03	1.75e-03	1.388e-03	1.374e-03	5.790e-05	5.6e-04	5.790e-05	5.790e-05
	0.0	4.758e-03	4.252e-03	4.858e-03	4.237e-03	9.837e-04	1.10e-03	9.877e-04	9.552e-04	4.144e-05	3.3e-04	4.144e-05	4.142e-05
	0.5	2.802e-03	2.389e-03	2.873e-03	2.339e-03	5.863e-04	6.25e-04	5.891e-04	5.599e-04	2.496e-05	1.8e-04	2.496e-05	2.494e-05
	0.9	1.290e-03	1.150e-03	1.319e-03	1.148e-03	2.714e-04	3.14e-04	2.726e-04	2.638e-04	1.175e-05	1.0e-04	1.175e-05	1.175e-05
5	−0.9	2.387e-02	2.866e-02	2.367e-02	2.347e-02	4.757e-03	7.56e-03	4.749e-03	4.736e-03	1.964e-04	2.4e-03	1.964e-04	1.963e-04
	−0.5	1.900e-02	2.028e-02	1.901e-02	1.900e-02	3.798e-03	4.89e-03	3.798e-03	3.798e-03	1.568e-04	1.2e-03	1.568e-04	1.567e-04
	0.0	1.297e-02	1.261e-02	1.306e-02	1.261e-02	2.599e-03	2.74e-03	2.603e-03	2.576e-03	1.072e-04	4.0e-04	1.072e-04	1.072e-04
	0.5	6.978e-03	6.329e-03	7.061e-03	5.459e-03	1.400e-03	1.26e-03	1.403e-03	1.179e-03	5.777e-05	6.5e-05	5.777e-05	5.703e-05
	0.9	2.184e-03	1.952e-03	2.218e-03	1.842e-03	4.389e-04	4.18e-04	4.403e-04	4.147e-04	1.825e-05	5.8e-05	1.825e-05	1.820e-05
4	−0.9	1.530e+06	–	1.903e+06	6.423e+05	1.110e+02	–	1.169e+03	1.110e+02	1.502	–	5.340	1.502
	−0.5	7.204e+06	–	8.906e+06	6.503e+05	6.978e+01	–	3.318e+04	6.978e+01	1.449	–	3.157	1.449
	0.0	4.367e+05	–	5.604e+05	3.481e+05	7.209e+01	–	3.853e+02	7.205e+01	1.386	–	2.437	1.386
	0.5	9.913e+06	–	1.224e+07	3.920e+05	1.074e+02	–	7.260e+02	1.074e+02	1.324	–	2.116	1.324
	0.9	2.050e+06	–	2.534e+06	2.491e+05	8.145e+01	–	1.287e+04	8.143e+01	1.273	–	2.789	1.272
6	−0.9	1.840e+06	–	2.287e+06	7.366e+05	1.337e+02	–	1.289e+03	1.337e+02	1.776	–	5.775	1.776
	−0.5	7.596e+06	–	9.389e+06	5.888e+05	7.816e+01	–	2.991e+04	7.816e+01	1.612	–	3.141	1.612
	0.0	4.178e+05	–	5.339e+05	3.246e+05	7.238e+01	–	5.572e+02	7.236e+01	1.411	–	2.613	1.410
	0.5	9.034e+06	–	1.116e+07	3.332e+05	9.641e+01	–	5.725e+02	9.641e+01	1.210	–	1.844	1.210
	0.9	1.612e+06	–	1.993e+06	1.953e+05	6.729e+01	–	9.126e+03	6.727e+01	1.049	–	2.242	1.049

Table E.2: $\widehat{\text{MSE}}(T_l, f(\boldsymbol{\mu}))$, $l = 1, 2, 3, \kappa_{min}$ for the Exponential Distribution.

PS	ρ	$n=10$: T_1	$n=10$: T_2	$n=10$: T_3	$n=10$: $T_{\kappa_{min}}$	$n=100$: T_1	$n=100$: T_2	$n=100$: T_3	$n=100$: $T_{\kappa_{min}}$	$n=1200$: T_1	$n=1200$: T_2	$n=1200$: T_3	$n=1200$: $T_{\kappa_{min}}$
1	−0.9	3.792e-03	4.360e-03	3.774e-03	3.763e-03	7.512e-04	1.16e-03	7.507e-04	7.504e-04	3.110e-05	4.0e-04	3.110e-05	3.108e-05
	−0.5	2.991e-03	3.349e-03	2.980e-03	2.972e-03	5.931e-04	8.49e-04	5.927e-04	5.925e-04	2.455e-05	2.5e-04	2.455e-05	2.454e-05
	0.0	1.992e-03	2.156e-03	1.987e-03	1.983e-03	3.954e-04	5.09e-04	3.952e-04	3.951e-04	1.637e-05	1.1e-04	1.637e-05	1.636e-05
	0.5	9.955e-04	1.040e-03	9.939e-04	9.927e-04	1.977e-04	2.26e-04	1.977e-04	1.976e-04	8.186e-06	3.3e-05	8.186e-06	8.181e-06
	0.9	1.990e-04	2.023e-04	1.989e-04	1.987e-04	3.955e-05	4.10e-05	3.955e-05	3.954e-05	1.637e-06	2.6e-06	1.637e-06	1.637e-06
2	−0.9	1.536e-02	1.753e-02	1.528e-02	1.524e-02	3.044e-03	4.61e-03	3.041e-03	3.040e-03	1.261e-04	1.5e-03	1.261e-04	1.260e-04
	−0.5	1.293e-02	1.406e-02	1.289e-02	1.286e-02	2.567e-03	3.30e-03	2.566e-03	2.565e-03	1.065e-04	7.4e-04	1.065e-04	1.064e-04
	0.0	9.926e-03	1.030e-02	9.907e-03	9.872e-03	1.972e-03	2.12e-03	1.971e-03	1.970e-03	8.191e-05	1.8e-04	8.192e-05	8.187e-05
	0.5	6.948e-03	7.179e-03	6.933e-03	6.886e-03	1.378e-03	1.44e-03	1.377e-03	1.372e-03	5.734e-05	8.8e-05	5.733e-05	5.720e-05
	0.9	4.585e-03	5.132e-03	4.562e-03	4.537e-03	9.030e-04	1.25e-03	9.021e-04	9.009e-04	3.764e-05	3.5e-04	3.763e-05	3.759e-05
3	−0.9	8.586e-03	9.903e-03	8.544e-03	8.518e-03	1.700e-03	2.66e-03	1.699e-03	1.698e-03	7.036e-05	9.3e-04	7.037e-05	7.033e-05
	−0.5	6.990e-03	7.940e-03	6.960e-03	6.943e-03	1.384e-03	2.08e-03	1.383e-03	1.383e-03	5.725e-05	6.9e-04	5.726e-05	5.723e-05
	0.0	4.994e-03	5.578e-03	4.977e-03	4.968e-03	9.882e-04	1.43e-03	9.877e-04	9.875e-04	4.088e-05	4.5e-04	4.088e-05	4.087e-05
	0.5	2.999e-03	3.322e-03	2.989e-03	2.984e-03	5.924e-04	8.41e-04	5.921e-04	5.920e-04	2.451e-05	2.5e-04	2.451e-05	2.451e-05
	0.9	1.401e-03	1.591e-03	1.394e-03	1.390e-03	2.755e-04	4.13e-04	2.753e-04	2.752e-04	1.143e-05	1.3e-04	1.143e-05	1.143e-05
4	−0.9	2.374e-02	2.723e-02	2.363e-02	2.356e-02	4.705e-03	7.23e-03	4.701e-03	4.699e-03	1.948e-04	2.4e-03	1.948e-04	1.947e-04
	−0.5	1.893e-02	2.095e-02	1.886e-02	1.881e-02	3.755e-03	5.16e-03	3.753e-03	3.752e-03	1.555e-04	1.4e-03	1.556e-04	1.555e-04
	0.0	1.293e-02	1.371e-02	1.290e-02	1.287e-02	2.568e-03	3.04e-03	2.567e-03	2.566e-03	1.065e-04	5.0e-04	1.065e-04	1.064e-04
	0.5	6.953e-03	7.121e-03	6.943e-03	6.921e-03	1.382e-03	1.43e-03	1.382e-03	1.380e-03	5.734e-05	8.1e-05	5.734e-05	5.732e-05
	0.9	2.188e-03	2.312e-03	2.182e-03	2.172e-03	4.332e-04	5.01e-04	4.329e-04	4.323e-04	1.802e-05	7.2e-05	1.802e-05	1.799e-05
5	−0.9	2.745e+06	–	3.400e+06	6.497e+05	1.093e+02	–	9.676e+02	1.093e+02	1.461	–	2.317	1.460
	−0.5	1.645e+08	–	2.031e+08	9.263e+05	1.454e+02	–	1.624e+04	1.454e+02	1.411	–	1.495	1.411
	0.0	3.877e+05	–	4.788e+05	2.683e+04	3.117e+03	–	3.303e+03	3.040e+03	1.347	–	1.467e+02	1.347
	0.5	7.310e+05	–	9.112e+05	3.607e+05	1.820e+02	–	3.323e+02	1.819e+02	1.283	–	6.849	1.282
	0.9	1.495e+10	–	1.846e+10	4.150e+06	1.081e+02	–	1.063e+04	1.081e+02	1.231	–	1.377	1.230
6	−0.9	3.378e+06	–	4.182e+06	7.247e+05	1.289e+02	–	1.015e+03	1.289e+02	1.728	–	2.857	1.727
	−0.5	1.860e+08	–	2.296e+08	1.036e+06	1.583e+02	–	2.141e+04	1.583e+02	1.571	–	1.626	1.571
	0.0	3.946e+05	–	4.875e+05	3.362e+04	3.106e+03	–	3.294e+03	3.032e+03	1.373	–	8.989e+01	1.373
	0.5	7.069e+05	–	8.791e+05	3.007e+05	1.683e+02	–	3.250e+02	1.682e+02	1.174	–	4.275	1.174
	0.9	1.174e+10	–	1.449e+10	4.044e+06	8.873e+01	–	1.140e+04	8.873e+01	1.015	–	1.149	1.014

Table E.3: $\widehat{\text{MSE}}(T_l, f(\boldsymbol{\mu}))$, $l = 1, 2, 3, \kappa_{min}$ for the Uniform Distribution.

E.2 $\widehat{\textbf{MSE}}(T_l, f(\boldsymbol{\mu}))$, $l = 1, 2, 3, \kappa_{min}, ML, MVUE_c, Shaban, Shaban1$ under Lognormal Distribution Assumption

n	PS	T_1	T_2	T_3	$T_{\kappa_{min}}$	$\hat{\varrho}_{ML}$	$\hat{\varrho}_{MVUE_c}$	$\hat{\varrho}_{Shaban}$	$\hat{\varrho}_{Shaban1}$
10	1	2.0252e-05	2.0265e-05	2.0252e-05	2.0252e-05	2.0251e-05	2.0250e-05	2.0250e-05	2.0344e-05
	2	6.7447e-03	6.7474e-03	6.7448e-03	6.7446e-03	6.7455e-03	6.7456e-03	6.7460e-03	6.7966e-03
	3	3.0111e-03	3.0121e-03	3.0112e-03	3.0110e-03	3.0104e-03	3.0104e-03	3.0105e-03	3.0259e-03
	4	2.0367e-03	2.1916e-03	2.0327e-03	2.0314e-03	2.0297e-03	2.0229e-03	2.0153e-03	3.0174e-03
	5	6.7031e-01	7.1945e-01	6.6923e-01	6.6899e-01	6.6808e-01	6.6645e-01	6.6516e-01	1.0124
	6	3.0080e-01	3.2264e-01	3.0040e-01	3.0036e-01	2.9898e-01	2.9806e-01	2.9712e-01	4.5194e-01
50	1	4.0281e-06	4.0335e-06	4.0282e-06	4.0276e-06	4.0276e-06	4.0277e-06	4.0280e-06	4.1379e-06
	2	1.3321e-03	1.3357e-03	1.3321e-03	1.3321e-03	1.3322e-03	1.3322e-03	1.3322e-03	1.3641e-03
	3	5.9189e-04	5.9225e-04	5.9191e-04	5.9167e-04	5.9192e-04	5.9195e-04	5.9200e-04	6.1001e-04
	4	4.0029e-04	5.0163e-04	4.0032e-04	4.0029e-04	3.9826e-04	3.9821e-04	3.9836e-04	1.3261e-03
	5	1.2992e-01	1.6830e-01	1.2985e-01	1.2980e-01	1.2928e-01	1.2919e-01	1.2909e-01	4.2588e-01
	6	6.0359e-02	7.7387e-02	6.0330e-02	6.0309e-02	6.0073e-02	6.0033e-02	5.9988e-02	1.9205e-01
100	1	1.9308e-06	1.9377e-06	1.9308e-06	1.9305e-06	1.9313e-06	1.9314e-06	1.9315e-06	2.0389e-06
	2	6.6188e-04	6.6541e-04	6.6188e-04	6.6188e-04	6.6163e-04	6.6163e-04	6.6163e-04	6.9068e-04
	3	2.9904e-04	3.0118e-04	2.9904e-04	2.9898e-04	2.9902e-04	2.9901e-04	2.9900e-04	3.1105e-04
	4	2.0014e-04	2.9910e-04	2.0018e-04	2.0011e-04	1.9873e-04	1.9874e-04	1.9882e-04	1.1014e-03
	5	6.4842e-02	1.0112e-01	6.4817e-02	6.4784e-02	6.4611e-02	6.4578e-02	6.4530e-02	3.5181e-01
	6	3.0876e-02	4.6681e-02	3.0871e-02	3.0869e-02	3.0595e-02	3.0585e-02	3.0574e-02	1.6119e-01
1200	1	1.7325e-07	1.8317e-07	1.7325e-07	1.7325e-07	1.7331e-07	1.7331e-07	1.7331e-07	2.6405e-07
	2	5.5728e-05	5.8745e-05	5.5729e-05	5.5722e-05	5.5735e-05	5.5735e-05	5.5735e-05	8.6502e-05
	3	2.4673e-05	2.6347e-05	2.4672e-05	2.4667e-05	2.4674e-05	2.4674e-05	2.4674e-05	3.7535e-05
	4	1.6606e-05	1.1850e-04	1.6606e-05	1.6604e-05	1.6559e-05	1.6559e-05	1.6558e-05	8.9261e-04
	5	5.4597e-03	3.9299e-02	5.4593e-03	5.4574e-03	5.4529e-03	5.4524e-03	5.4516e-03	2.9279e-01
	6	2.4980e-03	1.7412e-02	2.4981e-03	2.4979e-03	2.4947e-03	2.4948e-03	2.4950e-03	1.3307e-01

Table E.4: $\widehat{\text{MSE}}(T_l, f(\boldsymbol{\mu}))$, $l = 1, 2, 3, \kappa_{min}, ML, MVUE_c, Shaban, Shaban1$. 1st and 2nd Situations

n	PS	T_1	T_2	T_3	$T_{\kappa_{min}}$	$\hat{\varrho}_{ML}$	$\hat{\varrho}_{Shaban1}$
10	1	1.0401e-03	1.0410e-03	1.0400e-03	1.0384e-03	1.0475e-03	1.0043e-03
	2	3.4194e-01	3.4238e-01	3.4190e-01	3.4076e-01	3.4435e-01	3.3065e-01
	3	1.5400e-01	1.5416e-01	1.5399e-01	1.5370e-01	1.5509e-01	1.4906e-01
	4	8.6135e-01	8.9085e-01	8.5858e-01	8.2304e-01	1.1848	1.5522e-01
	5	2.7583e+02	2.8611e+02	2.7487e+02	2.6241e+02	3.8144e+02	5.0146e+01
	6	1.2818e+02	1.3298e+02	1.2772e+02	1.2183e+02	1.8045e+02	2.2566e+01
50	1	2.2515e-04	2.2612e-04	2.2513e-04	2.1781e-04	2.2619e-04	1.9963e-04
	2	7.6019e-02	7.6386e-02	7.6012e-02	7.2828e-02	7.6398e-02	6.6664e-02
	3	3.4037e-02	3.4174e-02	3.4034e-02	3.3014e-02	3.4188e-02	3.0572e-02
	4	4.9625e-01	5.1926e-01	4.9580e-01	3.9504e-01	5.2766e-01	2.1322e-02
	5	1.6855e+02	1.7594e+02	1.6841e+02	1.3714e+02	1.7536e+02	7.0301
	6	7.4859e+01	7.8374e+01	7.4791e+01	5.9296e+01	7.8765e+01	3.2286
100	1	1.2359e-04	1.2458e-04	1.2358e-04	1.1157e-04	1.2410e-04	9.9261e-05
	2	4.2471e-02	4.2805e-02	4.2468e-02	3.8356e-02	4.2642e-02	3.4338e-02
	3	1.9214e-02	1.9376e-02	1.9212e-02	1.7149e-02	1.9291e-02	1.5335e-02
	4	4.5284e-01	4.7495e-01	4.5263e-01	2.9988e-01	4.6150e-01	1.0332e-02
	5	1.5196e+02	1.5952e+02	1.5189e+02	9.9304e+01	1.5510e+02	3.4519
	6	6.7348e+01	7.0723e+01	6.7315e+01	4.2828e+01	6.9307e+01	1.5860
1200	1	3.3416e-05	3.4421e-05	3.3415e-05	1.0774e-05	3.3458e-05	8.7881e-06
	2	1.0821e-02	1.1152e-02	1.0821e-02	3.4532e-03	1.0835e-02	2.7848e-03
	3	5.0257e-03	5.1774e-03	5.0256e-03	1.5771e-03	5.0317e-03	1.2694e-03
	4	4.1614e-01	4.3772e-01	4.1612e-01	5.3434e-02	4.1642e-01	8.6165e-04
	5	1.3685e+02	1.4391e+02	1.3684e+02	1.7621e+01	1.3709e+02	2.7827e-01
	6	6.1453e+01	6.4627e+01	6.1450e+01	8.2762	6.1576e+01	1.2451e-01

Table E.5: $\widehat{\text{MSE}}(T_l, f(\boldsymbol{\mu}))$, $l = 1, 2, 3, \kappa_{min}, ML, Shaban1$. 3rd Situation

E.3 $\widehat{\text{MSE}}(T_l, f(\boldsymbol{\mu}))$, $l = 1, 2, 3, \kappa_{min}, CROW$ under Gamma Distribution Assumption

n	PS	T_1	T_2	T_3	$T_{\kappa_{min}}$	T_{CROW}
	1	5.4649e-02	2.6341e-01	5.1172e-02	5.0884e-02	5.2830e-02
	2	3.4306e-02	1.0169e-01	3.3010e-02	3.2825e-02	3.1944e-02
10	3	1.0023	–	2.9952e+14	1.0020	7.7683e-01
	4	8.6697e-03	2.3191e-02	8.3396e-03	8.2602e-03	8.3480e-03
	5	4.9680e-03	2.0326e-02	4.6747e-03	4.6339e-03	4.6820e-03
	6	1.2237e+01	–	1.2622e+16	1.2237e+01	7.2037
	1	9.9743e-03	1.4287e-01	9.8204e-03	9.7689e-03	9.9039e-03
	2	6.8784e-03	5.3918e-02	6.8078e-03	6.7685e-03	6.7631e-03
50	3	9.4085e-02	–	5.6197e+11	9.4043e-02	8.8866e-02
	4	1.6281e-03	1.1181e-02	1.6128e-03	1.6033e-03	1.6151e-03
	5	9.6966e-04	9.9235e-03	9.5396e-04	9.4254e-04	9.5717e-04
	6	1.0952	–	1.2082e+16	1.0951	9.9572e-01
	1	5.0946e-03	1.2477e-01	5.0698e-03	5.0663e-03	5.0820e-03
	2	3.3194e-03	4.6331e-02	3.3046e-03	3.2980e-03	3.2993e-03
100	3	4.1108e-02	–	1.2609e+11	4.1108e-02	4.0021e-02
	4	8.3417e-04	9.6367e-03	8.3004e-04	8.2686e-04	8.3078e-04
	5	4.5745e-04	8.2031e-03	4.5474e-04	4.5348e-04	4.5514e-04
	6	5.1699e-01	–	2.9868e+13	5.1683e-01	4.9081e-01
	1	4.3806e-04	1.1165e-01	4.3821e-04	4.3805e-04	4.3809e-04
	2	2.6680e-04	4.0706e-02	2.6668e-04	2.6661e-04	2.6662e-04
1200	3	3.3489e-03	–	1.2263e+08	3.3487e-03	3.3395e-03
	4	6.9738e-05	8.0386e-03	6.9718e-05	6.9709e-05	6.9720e-05
	5	3.9492e-05	7.0260e-03	3.9479e-05	3.9475e-05	3.9480e-05
	6	3.9929e-02	–	2.0423e+13	3.9921e-02	3.9748e-02

Table E.6: $\widehat{\text{MSE}}(T_l, f(\boldsymbol{\mu}))$, $l = 1, 2, 3, \kappa_{min}, CROW$.

Appendix F

Approximated and Simulated Results for the Estimation of the Inverse Mean

F.1 Comparison of Estimators T_1 and $T_{\kappa_{min}}$. Normal Distribution

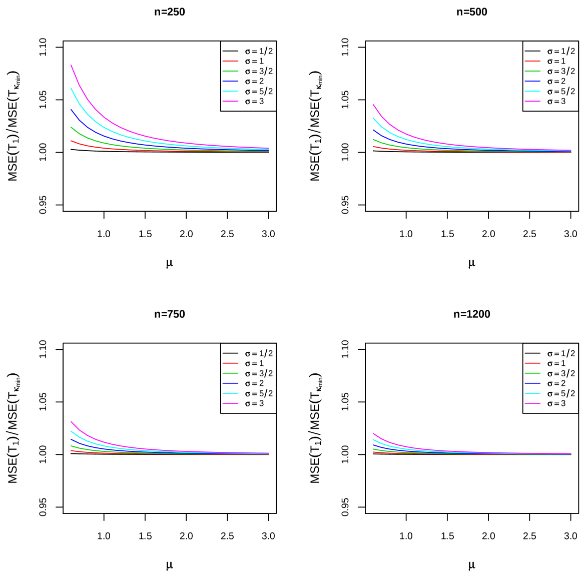

Figure F.1: $\mathrm{MSE}(T_1^{Poly}, f^{Poly}(\mu))/\mathrm{MSE}(T_{\kappa_{min}}^{Poly}, f^{Poly}(\mu))$.

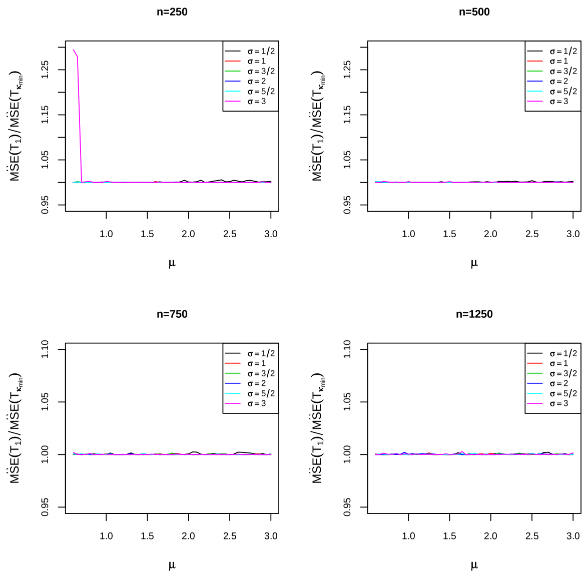

Figure F.2: $\widehat{\mathrm{MSE}}(T_1, f(\mu))/\widehat{\mathrm{MSE}}(T_{\kappa_{min}}, f(\mu))$.

F.2 Comparison of Estimators T_1 and T_{Sriv}. Normal Distribution

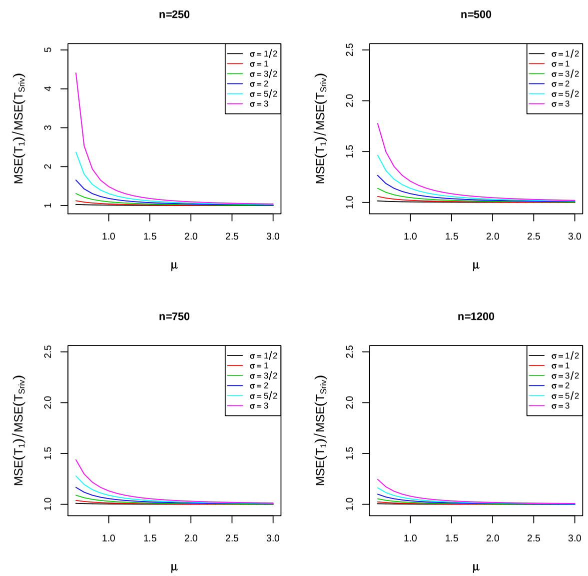

Figure F.3: $\text{MSE}(T_1^{Poly}, f^{Poly}(\mu))/\text{MSE}(T_{Sriv}^{Poly}, f^{Poly}(\mu))$.

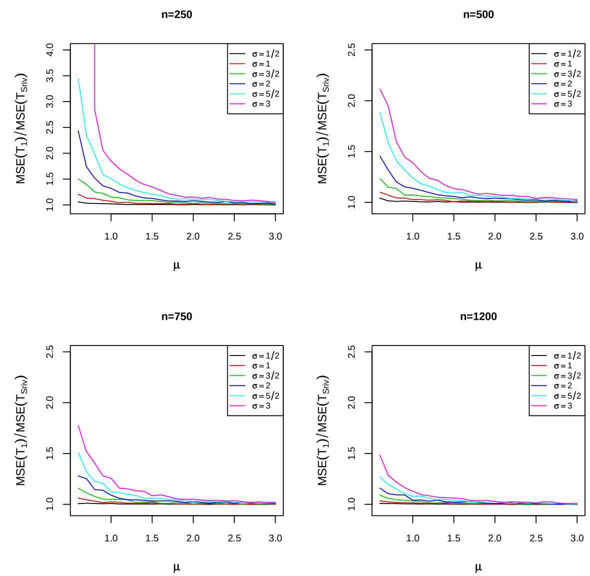

Figure F.4: $\widehat{\text{MSE}}(T_1, f(\mu))/\widehat{\text{MSE}}(T_{Sriv}, f(\mu))$.

F.3 Comparison of Estimators T_1 and T_2. Normal Distribution

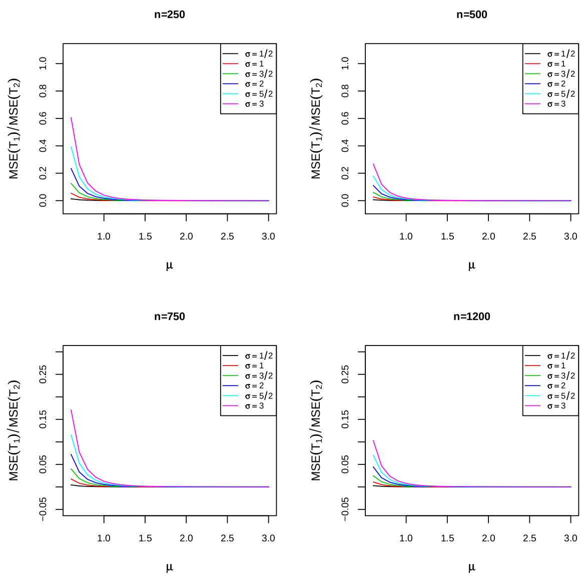

Figure F.5: $\mathrm{MSE}(T_1^{Poly}, f^{Poly}(\mu))/\mathrm{MSE}(T_2^{Poly}, f^{Poly}(\mu))$.

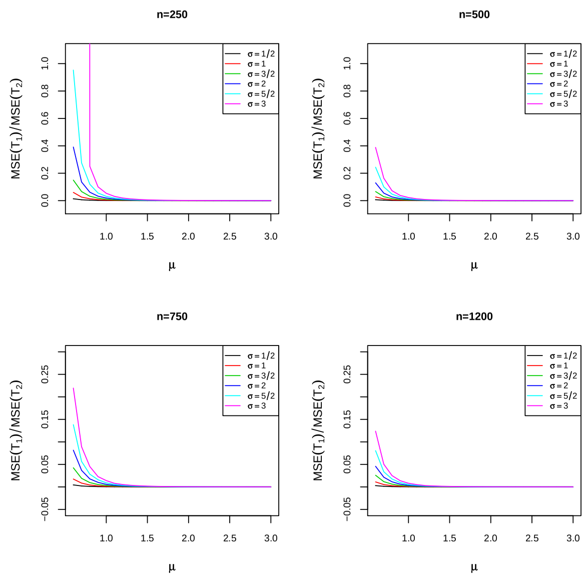

Figure F.6: $\widehat{\mathrm{MSE}}(T_1, f(\mu))/\widehat{\mathrm{MSE}}(T_2, f(\mu))$.

F.4 Comparison of Estimators $T_{\kappa_{min}}$ and T_{Sriv}. Normal Distribution

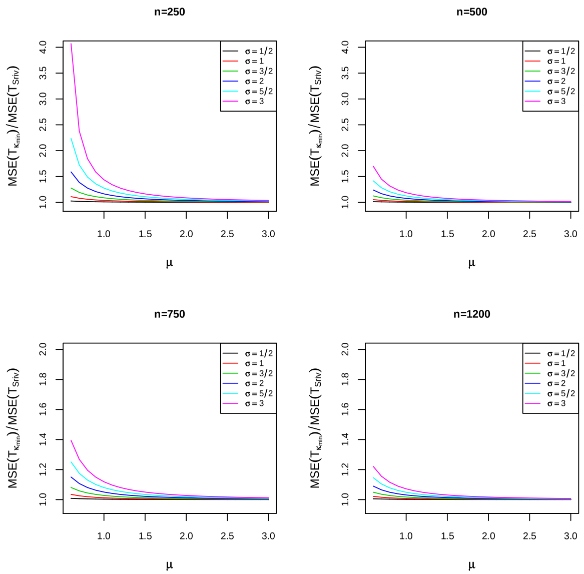

Figure F.7: $\mathrm{MSE}(T^{Poly}_{\kappa_{min}}, f^{Poly}(\mu))/\mathrm{MSE}(T^{Poly}_{Sriv}, f^{Poly}(\mu))$.

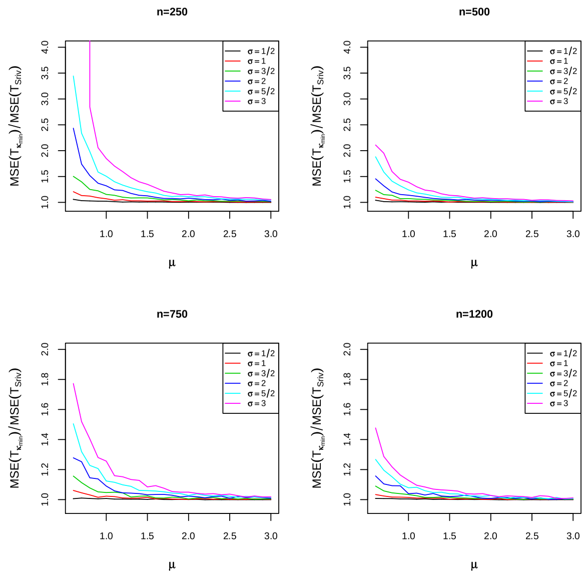

Figure F.8: $\widehat{\mathrm{MSE}}(T_{\kappa_{min}}, f(\mu))/\widehat{\mathrm{MSE}}(T_{Sriv}, f(\mu))$.

F.5 Comparison of Estimators T_1, T_3, $T_{\kappa_{min}}$ and T_{Sriv}. Exponential Distribution

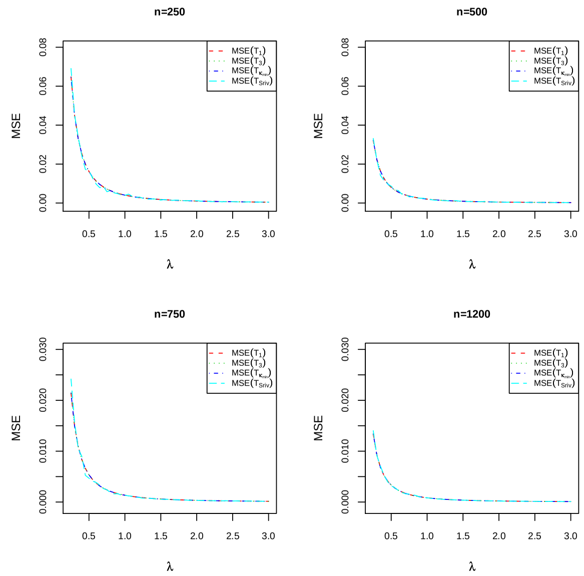

Figure F.9: $\mathrm{MSE}(T_l^{Poly}, f^{Poly}(\mu)), l = 1, 3, \kappa_{min}, Sriv.$

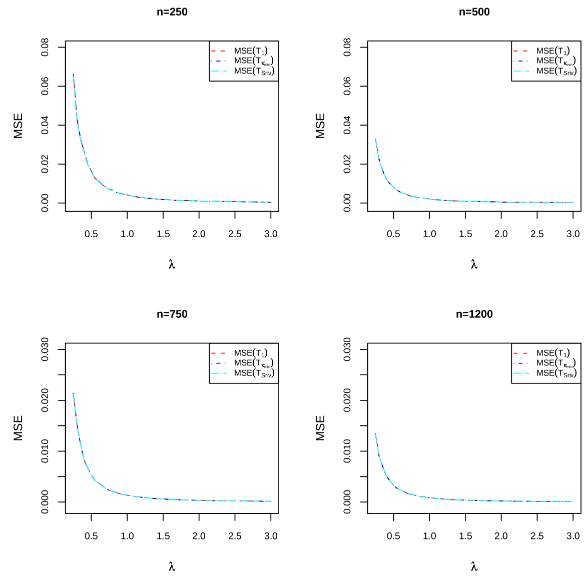

Figure F.10: $\widehat{\mathrm{MSE}}(T_l, f(\mu)), l = 1, 3, \kappa_{min}, Sriv.$

F.6 Comparison of Estimators T_1, T_3, $T_{\kappa_{min}}$ and T_{Sriv}. Uniform Distribution

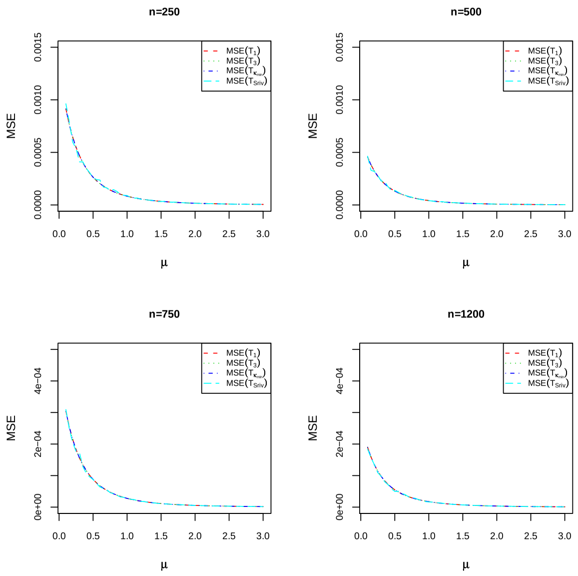

Figure F.11: $\mathrm{MSE}(T_l^{Poly}, f^{Poly}(\mu)), l = 1, 3, \kappa_{min}, Sriv.$

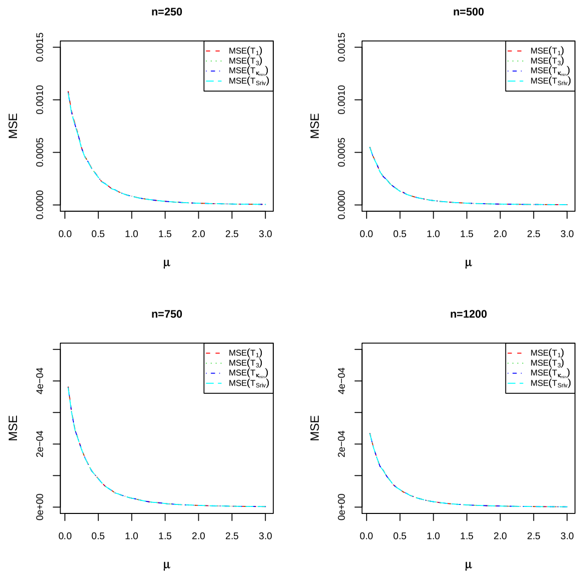

Figure F.12: $\widehat{\mathrm{MSE}}(T_l, f(\mu)), l = 1, 3, \kappa_{min}, Sriv.$

F.7 Estimator T_2. Exponential and Uniform Distribution

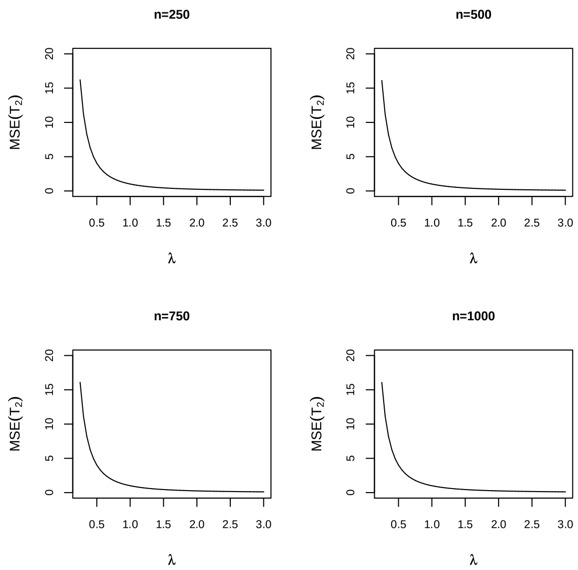

Figure F.13: MSE($T_2^{Poly}, f^{Poly}(\mu)$). Exponential Distribution

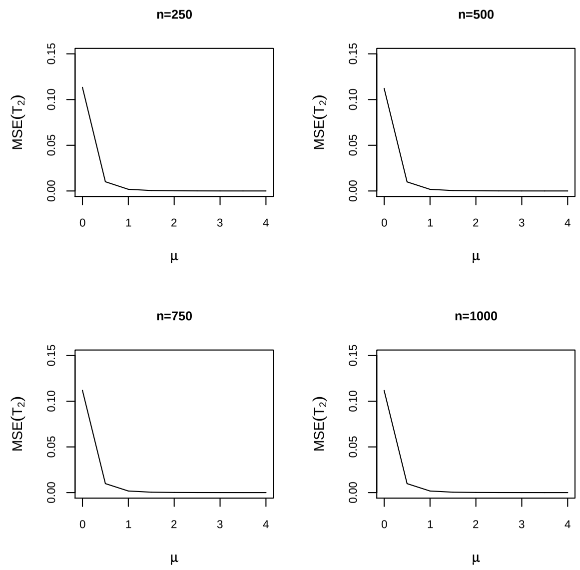

Figure F.14: MSE($T_2^{Poly}, f^{Poly}(\mu)$). Uniform Distribution

Appendix G

Approximated and Simulated Results for the Estimation of the Odds in Favour of an Event

		MSE			
PS	n	T_1^{Poly}	T_2^{Poly}	T_3^{Poly}	$T_{\kappa_{min}}^{Poly}$
	250	7.217e-05	5.461e-04	7.162e-05	5.786e-05
1	500	4.622e-05	6.072e-04	4.604e-05	4.209e-05
	750	3.554e-05	6.956e-04	3.545e-05	3.361e-05
	1200	1.944e-05	5.162e-04	1.941e-05	1.869e-05
	250	7.829e-04	1.649e-02	7.781e-04	7.558e-04
2	500	3.345e-04	1.237e-02	3.334e-04	3.279e-04
	750	2.568e-04	1.469e-02	2.562e-04	2.537e-04
	1200	1.491e-04	1.300e-02	1.489e-04	1.480e-04
	250	1.627e-01	4.616	1.627e-01	1.626e-01
3	500	7.755e-02	4.350	7.753e-02	7.753e-02
	750	4.493e-02	3.813	4.494e-02	4.493e-02
	1200	3.043e-02	4.081	3.044e-02	3.043e-02
	250	1.168e+03	7.845e+03	1.172e+03	1.141e+03
4	500	3.504e+02	5.039e+03	3.510e+02	3.468e+02
	750	4.641e+02	8.727e+03	4.646e+02	4.602e+02
	1200	2.404e+02	7.442e+03	2.406e+02	2.392e+02

		$\widehat{\text{MSE}}$			
PS	n	T_1	T_2	T_3	$T_{\kappa_{min}}$
	250	2.126e-05	2.140e-05	2.125e-05	2.054e-05
1	500	1.036e-05	1.039e-05	1.036e-05	1.023e-05
	750	7.018e-06	7.032e-06	7.018e-06	6.954e-06
	1200	4.414e-06	4.419e-06	4.414e-06	4.388e-06
	250	1.319e-04	1.330e-04	1.319e-04	1.311e-04
2	500	6.747e-05	6.786e-05	6.747e-05	6.694e-05
	750	4.608e-05	4.625e-05	4.608e-05	4.591e-05
	1200	2.829e-05	2.835e-05	2.829e-05	2.823e-05
	250	3.994e-03	4.147e-03	3.994e-03	3.960e-03
3	500	2.030e-03	2.067e-03	2.030e-03	2.023e-03
	750	1.353e-03	1.366e-03	1.353e-03	1.351e-03
	1200	8.576e-04	8.639e-04	8.576e-04	8.563e-04
	250	9.735e-01	1.318	9.728e-01	9.312e-01
4	500	4.857e-01	5.673e-01	4.856e-01	4.704e-01
	750	3.089e-01	3.422e-01	3.088e-01	3.031e-01
	1200	1.909e-01	2.030e-01	1.909e-01	1.889e-01

Table G.1: $\text{MSE}(T_l^{Poly}, f^{Poly}(\mu)), l = 1, 2, 3, \kappa_{min}$ and $\widehat{\text{MSE}}(T_l, f(\mu)),\ l = 1, 2, 3, \kappa_{min}$.

Ehrenwörtliche Erklärung

Hiermit versichere ich, die vorliegende Doktorarbeit unter der Betreuung von Prof. Dr. Trenkler nur mit den angegebenen Hilfsmitteln selbständig angefertigt zu haben.

Dortmund, den 31.08.2009 Ana Moya